

Disappointment about test bans

Last week's nuclear test beneath the Nevada desert probably spells the resumption of Soviet tests but need not imply that the end of the world is at hand. But technical people must be more alert.

In the end, after a couple of postponements, the threatened US nuclear test beneath the Nevada desert took place last week, the Soviet Union's news agency TASS announced that the eight-month Soviet self-denial on nuclear testing would be ended and people were generally thrown into despondency that the opportunity of concluding a comprehensive test-ban treaty had once again been lost. Luckily, things are not that bad. Both of the major superpowers are probably more realistic than they have been for several years about the need for agreements on strategic arms and about the difficulties of securing them. The plain truth is that, for the past three years, since the painful resumption of negotiations, the comprehensive test ban has not been near the top of the shopping-list canvassed by the United States. The United States has come round to the belief that a deal on missiles of intermediate range in Europe would be worthwhile, that a deal on strategic missiles would be tolerable, but that, while the dream of an entirely novel strategic balance built around the Strategic Defense Initiative remains a gleam in people's eyes, there is no possibility of relinquishing the opportunity of testing novel warheads. There is also the more immediate need to test warheads for the Midgetman missile, the workable alternative to the unworkable MX. These circumstances may be as unpalatable as they are disappointing, but there is no pretending that they are unreal.

Those who would see success in strategic arms control must above all be patient. The handful of agreements won in the past quarter of a century are a sufficient proof of that. So, in a different spirit, is the first-hand account by John Wright of how arms control negotiations are conducted published earlier this year (*Nature* 319, 275; 1986). But it is also well known that a large part of the process by which one partner brings another to the point of signing a treaty that has been agreed in outline consists of public shame. Thus the partial test ban concluded in 1963 came about largely because the two then superpowers (China had not exploded its first weapon) were persuaded by other governments that the dangers of radioactive fallout could not be indefinitely ignored. On this occasion, the Soviet Union has applied pressure of the same kind by offering a test ban, backed up by a voluntary moratorium, but to no avail. The result has been to strengthen suspicions that the United States may not be entirely serious about arms control. The fact that this pressure may have failed on this occasion is neither here nor there, however. What matters is merely that these events should not become impediments to an agreement when the time is ripe. The Soviet Union, to its credit, has said that the question of a test ban may be taken up again when people are so minded; let us hope it has not changed its mind by then.

Meanwhile, it would make sense that people in the West should prepare themselves more deliberately than has been their recent custom for the difficulties that lie ahead. Empiricism must be the order of the day, and the political irreversibility of arms control measures the best means towards more distant agreements. The plain fact, amply demonstrated by recent experience, is that governments that find themselves bound by arms control agreements, even those entered into by their predecessors, are almost always unable to shake loose from

them. This is the spirit in which, last year, the United States chose to lay up a nuclear submarine rather than to breach the unratified SALT II agreement on strategic arms. Similarly, in spite of temptations in the other direction occasioned by the Strategic Defense Initiative (SDI), the United States has not so far thought of tearing up the Anti-Ballistic Missile Treaty but has, instead, sought to preserve appearances by a reinterpretation of the legal ambiguities of the document. (If SDI lasts, the treaty will probably have to be renegotiated, which may be a restatement of the United States position that deployment would turn on disclosure to the Soviet Union.) So the process of arms control is a kind of ratchet. It may be slow, too slow for comfort, but it seems mostly to work in a beneficent direction.)

That does not argue for complacency. Indeed, there are the strongest possible reasons why the technical community should be more active than it has been in clarifying the basis on which agreement may be possible. The comprehensive test ban is a case in point. The abortive negotiations that ended in 1980 had already demonstrated that the technical ambiguities of the seismic detection of distant explosions could be circumvented by remotely sited seismometers. Since then, there has been considerable improvement in the techniques of remote detection. Now there is a rich vein of inventiveness about the ways in which particular objections to a comprehensive test ban might be accommodated (simply by postdating the agreement). There is a host of ways in which other goals might be assured by simple devices such as these, which come more naturally to technical people than to the politicians in which these momentous issues are constitutionally entrusted. Something, the technical community may conclude, should urgently be done. □

Perceptions of Europe

Europe is repeatedly offended by what it sees as US aggression in commercial matters.

FOR the past several weeks, electronics companies in Europe have been in a state bordering on paranoia over the prospect that the governments of the United States and of Japan would come to a commercial arrangement about the prices to be charged for microchips, initially in the United States but probably throughout the international market for these devices. The issues are simple (see p. 567). For several years, but insistently in the past several months, US manufacturers have been complaining to Washington about the prices at which Japanese manufacturers are able to sell microchips in the United States, which are substantially lower than those at which it is possible to manufacture them domestically. The point has now been reached at which the complainants' case has been declared superficially substantial; until the complaint is formally investigated and adjudicated, those who import Japanese chips into the United States are being required to make matching deposits of money commensurate with the penalties they may eventually have to pay if Japanese manufacturers are held to be dumping their goods on the US market at unfairly low prices. Whatever the eventual decision, which could be much delayed, the immediate



consequence has been to increase the price of semiconductors in the United States, to the relief of domestic manufacturers, many of whom may prefer the state of affairs thus created to that which would arise if the suits against Japan were decided and found to be invalid. What European manufacturers fear is that, in such circumstances, Japanese manufacturers will be forced into a deal with their counterparts in the United States whose effect will be to increase the price of microchips elsewhere, to the general disbenefit. Cartel is the dirty word that people use to describe this apprehension.

Plainly, it is not at present justifiable for people in Europe to describe the negotiations between US and Japanese manufacturers, or the respective governments, in language of this kind. Nothing like a formal cartel agreement is within sight of negotiation. The more substantial threat to the rest of the world is that Japan and the United States between them control such a large proportion of the world's present production of microchips that the mere conduct of negotiations about price might encourage the manufacturers concerned to slip into understandings about the price of their products that other people would be required to pay through the nose for the benefits thereof. Europeans, and European industries in particular, have good cause to worry on this score, for there have in the past several years been too many occasions on which the commercial interests of the United States have been defended, not always successfully, beneath the umbrella of strategic security. The natural gas pipeline from the Soviet Union to the West is the example that most conspicuously stands out. But the remedy is in European hands. European companies are skilled at designing chips, but are not especially good at manufacturing them in bulk; the threat they perceive from the United States and Japan will be more easily countered if the two concerned should conspire artificially to enhance the prices of their products. Artificially high prices on the international market might even provide the breathing space in which a European industry might gather its strength and resources. In short, there is no case for defensive paranoia.

Even so, Europe does have a just but lesser cause for complaint. The negotiations between the United States and Japan on trade have been conducted (as is inevitable) between the world's economic superpowers in virtual indifference to how the rest of the world may be affected by their mutual relationships. Both to the United States and Japan, Europe has become a consumer market and one that seems to be in an old-fashioned way unreasonably protectionist, of its agriculture in particular. Here again, the remedy is in Europe's hands, but will not be easily attained. Putting real flesh and blood on the skeleton of the common market promised by the Treaty of Rome lies still a long way off. Ironically, the most common misunderstanding between the United States and Europe is that the former behaves as if the former were a reality already, which alone can explain why Washington expects Europe to be capable of a single reaction on matters such as, say, Libya, while Europe still relishes its interesting but sapping diversity. □

More apartheid ironies

A conference at Berkeley puts recent events in Southampton in context.

THE decision by the organizers of the World Archaeological Congress to be held in Southampton, England, in September to exclude participants from South Africa and Namibia has sharply divided the community of archaeologists, but needlessly. A conference being held this week at Berkeley, California, seems to show that when rhetoric and bombast are kept to a minimum, academic discussion can be conducted without rancour. The five-day conference at Berkeley under the title "The longest record: the human career in Africa", was organized by Jack Harris of the University of Wisconsin and the late G. Ll. Isaac of Harvard University to honour Professor J. Desmond Clark,

retiring this year from the University of California. Professor Clark has left an indelible mark on the study of African prehistory since the beginning of his professional life in the 1930s. He is respected not only for his academic contributions, but also for his diplomatic skill in maintaining research projects despite political obstacles in Africa.

The Berkeley conference seems destined to achieve precisely the goal of true international participation that prompted the organizers at Southampton to exclude South African researchers. The expectation that representatives from black African nations would boycott a conference attended by South Africans has failed to materialize at Berkeley. A mailing sent to all pre-registrants for the conference clearly listed participants from South Africa, Kenya, Malawi, Nigeria, Sudan, Tanzania and Zambia, as well as the United States, Canada and Europe. Not only has a boycott failed to materialize, the ranks of the conference have been swollen; 240 registrants paid the modest \$25 fee to attend. Formal sessions at the conference were scheduled to start at 8 a.m. and continue until 10 p.m. to accommodate the 160 papers submitted. Why such worldwide enthusiasm? Clearly Professor Clark has captured tremendous respect and affection from the people with whom he has worked. The efforts of organizers Isaac (before his untimely death last September) and Harris to include all aspects of early human activity in Africa have had an effect. By including a one-day symposium on rock art, Harris was able to coax a \$10,000 grant out of the financially strapped National Endowment for the Humanities. It is also the case that there has not been a good conference on early hominids in Africa since the Pan African Congress in 1977 in Kenya. There may also be other explanations of this enthusiasm, but it would seem that the conference has shown that a large segment of the prehistory community regards academic meetings as the wrong forum for making political statements.

The Berkeley conference has also shown what should in any case have been apparent, that successful meetings can take place even among communities sensitized to political issues. It can hardly be argued that students at Berkeley are significantly less interested in abolishing apartheid in South Africa than their counterparts in Southampton. Early this month, 87 students were arrested on the University of California campus in violent clashes with police over "shantytowns" erected on campus to call attention to the plight of South African blacks. There has been no disruption of the anthropology meeting. Not that the organizers were unaware that disruptions were possible. They merely chose not to call attention to their activities. Since the meeting coincides with the spring break on the Berkeley campus, student protests may have been blunted by hedonism, on which Southampton could also have counted. The issue of South African participation in international conferences should be an issue decided by individuals, not scientific bodies. It is one thing for a government, presumably acting on behalf of its people, to decide to isolate one country from the rest of the world community. But professionally competent though they may be, anthropologists and archaeologists should not be setting political agendas for international relations. If the moral issue of excluding South African scientists from scientific meetings were as clear-cut as the organizers at Southampton seem to believe, there would be no issue at all. Since the International Union of Prehistoric and Protohistoric Sciences (IUPPS) withdrew its support from the Southampton Congress, it has been clear that matters are not that cut and dried. One of life's little ironies of which the Southampton organizers will now be aware is that one of the participants at the Berkeley meeting is now also a member of the Southampton organizing committee. Another is that Harris suggested that Peter Ucko, Southampton organizing secretary, be a reviewer of the grant application to the National Endowment for the Humanities; the grant was used in part to pay travel expenses of South African participants. One wonders what Ucko's comments would have been. □

US space sciences

Martian base planned within forty years

Washington

DEVELOPMENT of new space vehicles to allow exploitation of lunar resources and place a manned base on Mars before the year 2025 is recommended in a report to be published shortly by the US National Commission on Space. A final draft of the commission's proposals, which have now been sent to the White House, describes an ambitious programme for manned and unmanned exploration of the inner Solar System supported by spaceports in low Earth and lunar orbits; manned bases would employ self-sustaining life support systems and use on-site materials. The first lunar outpost would be operational within 20 years.

The commission, appointed by President Reagan in March 1985 to formulate a long-term civilian space agenda, believes that manned exploration, prospecting and settlement of the near Solar System is feasible if the United States (and in particular the National Aeronautics and Space Administration, NASA) improves long-term planning. Establishing manned lunar bases will, the commission believes, spur development of technology that could later be adapted for use on Mars.

The commission recommends that the proportion of NASA's budget devoted to its "technology base" be increased from the present 2 per cent to 6 per cent, and that space programme budget planning be shifted to a five-year cycle. Multi-year procurement contracts should replace year-by-year contracts, while budget approval should be only every two years. The cost of the proposed space programme would increase from \$7,300 million per year (NASA's current budget) to around \$40,000 million by 2025 (1986 dollars); the commission points out that if (as it predicts) the US economy sustains a 2.4 per cent annual growth rate, the proportion of US gross domestic product being spent on space would remain less than half of the peak during the Apollo programme for a landing on the Moon.

The commission endorses foreign collaboration in space exploration and suggests that the United States should lead coalitions of participating nations. But it says that US commercial interests should be respected and stresses that the private sector should play an important role, especially in designing launch vehicles. Cooperation with the Soviet Union should be subject to special controls to prevent transfer of technology that might damage military security.

Anticipating future political battles over access to space resources, the com-

mission argues against the creation of new international organizations or new treaties to regulate broadly the use of space resources, believing that the four United Nations treaties to which the United States is party provide an adequate framework. It argues against US ratification of the "Moon treaty" dealing with natural resources on other celestial bodies, arguing that international regimes such as that envisaged in the treaty could inhibit US enterprise in space.

The commission lays much stress on the importance of maintaining continuity of launching services and of reducing space transport costs. Three new phased transport systems will be needed within the next 15 years: a cargo transport to reduce the cost of placing payloads in low Earth orbit to \$200 per pound, a low-Earth orbit passenger transport and a transfer vehicle operating from near-Earth orbit that could move payloads and crews over lunar distances and be piloted remotely to land on the Moon.

The passenger transport could be either an advanced reusable rocket or an air-breathing aerospace plane (NASA and the Department of Defense recently announced plans to develop an aerospace plane that would utilize supersonic combustion ramjet (scramjet) engines). Transfer vehicles would use hydrogen/oxygen fuel; for long interplanetary journeys, electric ion beam or mass-driver propulsion would be needed.

A major basic research effort is recommended to support exploration of the planets and asteroids. The planned space station now being developed would be followed before the year 2000 by a spaceport in low-Earth orbit, and a variable-gravity laboratory shortly thereafter. The commission stresses that much prospecting of the Solar System will have to be carried out remotely at great distances, and so recommends the development of exploring vehicles with substantial artificial intelligence for sample/return missions. And much research is still needed on the effects of prolonged weightlessness and exposure to radiation in space. Nevertheless, the commission is optimistic; Moon rocks might be used for shielding. The first crew would arrive on Mars by about 2010; later, the commission foresees continuously-cycling spacecraft in an energy-minimizing trajectory between Earth and Mars. The commission asks the administrator of NASA to prepare a five-year plan in response to its report by the end of this year.

Tim Beardsley

AIDS research

French virus in the picture

Washington

THE announcement in last week's *Science* that one of the key early papers on the AIDS (acquired immune deficiency syndrome) virus from Dr Robert Gallo's group at the National Institutes of Health includes an erroneously labelled composite electron micrograph will do nothing to help Gallo's group in its battle with the Institut Pasteur of Paris over patent rights to AIDS blood tests. The error can only be described as acutely embarrassing: electron micrographs in a paper by Schüpbach *et al.* (*Science* 224, 503; 1984) that are identified as HTLV-III isolated by Gallo's group in fact show samples of lymphadenopathy associated virus (LAV) supplied to Gallo's group in 1983 by Luc Montagnier at Institut Pasteur. Montagnier's lawyers argue that the photographs provide evidence that Gallo in some way benefited by using the Pasteur virus to make his own patent claim.

According to Gallo, the mistake was made by a technician at Frederick Cancer Research Center, which performed electron microscopy for Gallo after September 1983, and was discovered only recently. The photographs in question were sent to Gallo's laboratory in December 1983. Gallo has long maintained that he never succeeded in achieving more than transient growth of LAV, and that the blood test he developed was based on one of his own isolates; nevertheless, when a good micrograph of budding AIDS virus particles was called for in the Schüpbach paper, the picture selected was of LAV.

Gallo says the scientific content of the paper is unaffected by the use of the wrong photograph and insists that his group can prove conclusively that they had other good electron micrographs of HTLV-III together with evidence of reverse transcriptase activity (and negative tests for HTLV-I and HTLV-II) dating from as early as February 1983, before Montagnier's first publication of a photograph of LAV. These photographs, which were taken at a different laboratory, are shortly to be submitted for publication in a recognized journal in an attempt to lay to rest the suggestion made by Montagnier's lawyers that Gallo used LAV simply because that was his best source. Gallo points out that photographing the virus supplied by Montagnier was normal scientific practice provided for in the agreement accompanying the exchange to use the virus for research. Gallo lists in a letter to *Science* explaining the error (232, 307; 1986) three other papers (two in the same issue of *Science*) that included correctly identified HTLV-III.

Tim Beardsley

UK medical research

Charity begins at home

THE dangers of relying on charities or industry to prop up British medical research in the face of inadequate government support are emphasized in two new reports written from different perspectives. One, published by the Association of Medical Research Charities (AMRC), warns that despite a growing income, the charities cannot be expected to assume responsibility for "strengthening the research base". The other, published last week by the Office of Health Economics (OHE), warns that government policies may drive pharmaceutical companies to pursue their research elsewhere than in Britain.

In two years time, predicts the OHE report, the spending of charities on medical research is likely to outstrip that of the Medical Research Council. The latest figures from the charities association show that £89 million was spent by their members directly on medical research in 1984 compared with £113 million by MRC. But OHE predicts that the charities will contribute £133 million compared with £131 million from MRC in 1987-88. The latter figure will only increase in the unlikely event that the government alters its current economic policy. The predicted contribution from the charities incorporates the annual £15 million extra that Wellcome Trust expects as a result of its sale of 20 per cent of the Wellcome foundation early this year, and allows for an annual increase of 10 per cent in the sum available from other charities.

This is likely to be a modest forecast because in last month's budget, tax relief was granted on donations from companies of up to 3 per cent of their dividends and on contributions at source of up to £100 per annum from individual employees. This may rescue a situation in which a number of charities have had to dip into their reserves to provide continued support for essential research — 1985 was the first year in which the Imperial Cancer Research Fund spent more than it received and a year in which income increased by less than inflation according to figures released this week. But the shift away from government being the major provider of funds cannot continue "without inflicting permanent and lasting damage to the research enterprise in Britain", warns David Evered, director of the Ciba Foundation, in the AMRC report. Were the charities to divert some of their funds to bolster university departments deprived of government support, he adds, the public would probably respond by donating less money.

OHE supports this view, pointing out that whereas it is only because of charitable funds that medical research in Britain is still at a respectable level, they

cannot necessarily be relied upon, particularly for long-term support. (A clear exception is cancer research which accounts for very nearly half of the charities expenditure, much of it in long-term research.) Moreover, some diseases do not attract much in the way of donations from the public (in which respect Wellcome Trust can play a vital role), and charities cannot be expected to be responsible for training in research.

But of more concern to OHE, which was founded by the Association of the British Pharmaceutical Industry, is the future of the industry, which spent £490 million on research and development in 1984, in Britain. Repeating oft-heard complaints from an industry that claims to have discovered 5 of the 12 medicines that

had the biggest worldwide sales in 1984, OHE claims that moves by government to contain public expenditure threaten future investments by pharmaceutical companies. Particular concern is expressed about government measures designed to contain the cost of the National Health Service, including enforced price reductions and freezes for drugs, and the introduction of a policy, so far restricted in scope, that limits prescription on the National Health Service to a list of named drugs.

The overall result, says the OHE report, may be the rejection of Britain as a manufacturing and research base for the pharmaceutical industry, a conclusion for which some apparent support can be found in recent corporate decisions (see *Nature* 319, 529; 1986), even if there is a suspicion that other reasons may sometimes have been more important.

Peter Newmark

US intellectual property

Trade reprisals against pirates

Washington

PIRACY and counterfeiting has cost US companies between \$8,000 million and \$20,000 million in lost sales, according to Department of Commerce statistics, for which reason the Reagan administration is proposing legislation that will strengthen protection for US patents, copyrights and trademarks.

The strategy is aimed at those countries that either "flagrantly disregard" intellectual property rights or even encourage domestic industries to seek profits by copying US products. The administration said last week that the effects are to "distort" patterns of international trade and also to deprive innovators of the fruits of their labours. In the long run, the administration said, one result could be to stifle innovation and to discourage investment in new technology.

The proposed legislation is called the Intellectual Property Rights Improvement Bill and contains provisions that both protect US industry from international abuses and encourage continued research through modifications of licensing arrangements. The bill extends the rights of holders of process patents to include products made by the protected process.

Moreover, importing a product made under a patented process into the United States would be illegal. Since process patent holders may already prevent domestic manufacturers from using a patented process, this revision will primarily affect foreign manufacturers.

A widely supported component of the legislation would be modification of the procedure needed to block the sale of goods imported in violation of a property rights law. It would no longer be necessary

to show that the importation or sale will cause injury to a US industry, but simply that the importation infringes on an existing patent or trademark.

To encourage continued investment in research and development, the legislation offers several incentives to US companies. Patent terms for pesticides and veterinary products are extended to cover the time period used in the approval process. As things now stand, years of the patent period can be taken up waiting for permission from the appropriate federal agency to market a product. Such relief has already been provided for the pharmaceuticals industry.

The new legislation is just one part of a comprehensive strategy for protection of US intellectual property rights. The administration believes that the standards for protection in international conventions are too weak, especially in the patent area. In the long run, the United States is planning to seek development of international codes covering intellectual property. That process could begin at the next round of General Tariff and Trade Talks scheduled to start later this year. But in the meantime, bilateral "consultations" will be used to try to bring violating countries into line. Not only do many countries have inadequate laws covering intellectual property by US standards, but laws that do exist are inadequately enforced. The US Trade Representative has made it clear that violators face possible denial of trade benefits provided by the Generalized System of Preferences and the Caribbean Initiative. Resolving the issue of intellectual property protection is clearly a priority for the Reagan administration.

Joseph Palca

Bulgarian science

New organization for flexibility

BULGARIA'S Party Congress earlier this month approved "guidelines" for the economy that include a major departure from the typical socialist structure of the organization of science. The State Committee for Science and Technical Progress has been supplanted by a broader Council for Science, Culture and Education. And at a time when the Soviet Union is committed to reducing the number of institutes belonging neither to the academy network nor to the higher education sector, the Bulgarian party has recommended the establishment of research laboratories in major factories and enterprises.

The inclusion of "education" and "culture" in the title of the new council is meant to assist the "basic fundamental task" of the new body, namely the transfer of new research results and technologies to the production sector. This is a major problem for all the socialist bloc economies. The Bulgarian economic reform of the early 1980s recommended its own solution: research budgets were to be allocated on the basis of scientists' success in promoting their results in the production sector. Unlike other socialist countries, where universities and academy research institutes accept contract work proposed by industry and agriculture, Bulgarian scientists had to propose practical applications for their results in industry.

In practice, this has not proved effective. The president of the Bulgarian Academy of Sciences, Dr Angel Balevski, complained to the party congress that not only was scientists' time wasted when they had to act as salesmen for their work, but also the industrial managers who agreed to implement their results often claimed credit for the initial discovery. Balevski noted sourly that such behaviour "does nothing to boost the scientists' morale".

The new Council of Science, Culture and Education is meant to integrate research efforts. The usual pattern of research organization in socialist countries follows that established in the Soviet Union in the 1920s, namely, the Academy of Sciences and other academies (medicine, agriculture, pedagogic sciences and, in the Soviet Union, Czechoslovakia and Yugoslavia, the academies of individual republics), the universities and higher education institutions, and the "branch" institutes which come directly under the production ministries.

At the same time, the administration of science is divided between the State Committee, the academies and the Ministry of [Higher] Education. The pattern differs slightly from country to country; but in socialist countries, scientific research is effectively divided into three parts.

The Bulgarian "guidelines" aim to re-

duce the bureaucracy created by this division. They were first presented at a Central Committee plenary meeting in January in the unusual form of the personal thoughts of the party leader, Todor Zhivkov. As is not unusual in Bulgaria, the text of Zhivkov's speech was not published until much later, by which time several of the proposed changes had been implemented, including the abolition of several ministries and state committees and the creation of the three new councils with wide decision-making powers, whose decisions will have the validity of acts of the Council of Ministers.

This council, according to Zhivkov, is intended to "direct science towards education, education towards science and towards the people, both education and science towards culture, and so on". Like the other two councils, it is expected to "resolve questions, coordinate and ensure strategic leadership ... eliminate departmental barriers, and create opportunities for a more thorough and differentiated approach to the solution of problems."

This departure from the Soviet model, in a country often dubbed the "sixteenth republic" of the Soviet Union, has raised the speculation that it may be a trial run for changes to be implemented later throughout the socialist bloc. In fact, even in a small country like Bulgaria, it is not so easy to depart from the standard model. Although the new council will be responsible for financing the national science programme, it will not be "burdened" with the "direct leadership" of science, which will be left to the academies of science, agriculture and medicine and also the universities. Moreover, a new State Committee for Research and Technology (headed by Stoyan Markov, first deputy chairman of the Council of Ministers) has been established to deal with the practical side of technology transfer.

The party congress was more concerned with urging integration than discussing practical means. Already there is considerable overlap between the Bulgarian Academy of Sciences and Sofia's "Kliment of Ohrid" University, in the sense that a large number of scholars simultaneously hold posts in both. This integration, said the "theses" discussed and approved by the congress, must be "raised to a new level in conformity with contemporary criteria and requirements. Effective and efficient forms of integration must be applied in cooperation between the Bulgarian Academy of Sciences, the higher education institutions and the links of industrial science".

The higher education sector, however, is clearly not entirely happy with the new proposals. As the rector of Sofia Univer-

sity, Professor Mincho Semov, told the party congress, although his university plans to open a new technological centre in the near future and to be among the leading scientific centres of the world by the end of the century, it is, like all Bulgarian universities, hampered by an inequitable distribution of resources. Twenty-nine per cent of the country's scientists work in higher education, he said, but receive only five per cent of the science budget. Unless the proposed "integration" does something to even out this inequality, it is unlikely to attract the enthusiasm of Bulgarian scholars.

Vera Rich

UK computers

Sinclair eclipse

AFTER TWO rocky financial years, Britain's once proud computer company, Sinclair Research, has been forced by financial pressure to divest itself of its manufacturing and trading activities. In a major reorganization, the Sinclair group, which is the personal creation of Sir Clive Sinclair, has sold the rights to its trade name to the rival Amstrad group, which has grown into a substantial force in the computer market only in the past two years.

For its outlay of £7 million, Amstrad will also have the right to sell the more advanced business computer on which Sinclair has been working for some time. But Sir Clive Sinclair says that when this group has shed its other operations, his company will be free of the debt that has accumulated during the past few years, the better to pursue his interests in the development of computer technology as a consultant and contract researcher.

Sinclair's rise to international fame over the past decade was first based on the manufacture of cheap pocket calculators and wrist-watches based on microcircuits. His general distrust of the profits retained by the distributive trade has from the outset led him into unorthodox methods of marketing equipment by direct mail, to which end the group has been sustained by the loyalty of a hard core of its customers. Others, however, have been dismayed by the late delivery of promised equipment, and the difficulty of obtaining service.

After earlier financial troubles resolved only by the intervention of the British government's Enterprise Board, Sinclair introduced the first personal computer (called ZX80) to sell on the British market for less than £100. Five million of these machines have been sold since 1980. Of the outstanding projects in the Sinclair portfolio, one of the most radical is that for manufacturing large-scale integrated circuits on silicon wafers no less than 4 cm in diameter. This project is to be continued with the help of capital, said to amount to close on £50 million, from the venture capital arm of Barclay's Bank. □

Japanese education

Obsession with exam results

Tokyo

EARLIER this month the covers of many of Japan's national weeklies were not featuring the usual pictures of world leaders, or even film stars. Instead, beaming teenagers are seen literally jumping for joy (or being tossed into the air by a circle of their friends). The occasion? The *Sunday Mainichi* headline says it all: "Tokyo University-Kyoto University — the 6,000 successful candidates".

Given that entry to a famous university can guarantee a successful career for life, it is not wholly surprising that several best-selling weeklies see fit to print the names of every single successful candidate to the top two universities. But the obsession with educational results does not stop there. Inside the *Weekly Asahi* for example, there are twenty-five pages of examination results and analysis, including

pages and pages of statistics of the relative performance of the nation's high schools. This weekly has been building up to the Tokyo and Kyoto results with a series of detailed articles analysing results from other universities; before that, details of entrance examination papers and predictions of pass marks were carried. Several other magazines go into similar detail and they are not specialist educational magazines. The equivalent would be for *Time* or *Newsweek* to carry 30 or so pages of analysis of entrance requirements and results for Harvard and Yale.

The stages of the entrance examination are mirrored in the detailed articles that appear week by week. In one year, each student can apply to only one public university, for all hold their separate examinations on the same day. Early on, then, the emphasis is on predicting what the pass mark will be for any particular faculty so that the best bet can be selected. Students can measure their own ability easily enough at any of the companies offering mock examination services. Tiny shifts in probable pass marks are commented on in the press: a few points off an influential faculty's pass mark could put it within some anxious student's reach. Another faculty seems popular and its pass mark heads towards the sky.

At the end of January come the first real examinations, the part one "common examination" for all public universities. A great deal of attention is given to any new trends that may be apparent in the papers. Was it really reasonable to ask, for example, a question about the Mexican revolution when it occupies only a tiny corner of the textbook? "Good, orthodox questions" seems to be the ultimate compliment. With so much hanging on differences of a few points, nobody can bear to see even the tiniest change in the form of the game. And that is one reason why educational reform seems virtually impossible; however much it is called for, there is just too much at stake for those who have already begun training to permit any change.

The common entrance examination counts for only a part of the entrance score; the bulk of the marks come from the individual university examinations. So, after the part one is over there is still a chance to make up for a poor performance. Or is there? Series of articles will tell you. Using data collected from thousands of applicants' performances in part one, analysed according to previous trends, new predictions are made for the likely pass marks for all university faculties. From your own part one score, you can quickly look up your rank in the tables of probabilities that range from almost cer-

tain success to less than 20 per cent chance for recovery. For those in the latter categories, heroic efforts will be needed in the part two examinations.

The real flood of press coverage, however, does not begin until the results are announced a few weeks later. It is then that the entrants to every major university are listed according to high school — and the ability of schools to attract future students is determined. Even schools that send a single individual are in there. Closest attention goes to those who succeed in sending students to Tokyo and Kyoto Universities. For the past 20 years, the private school Nada, in Kobe, has dominated the rankings, always having taken the number one or two position. Nada is a unique private school. It succeeds in finishing the curriculum a year early so that an entire year can be given over to revision and mock entrance examinations. Other schools have tried to imitate it. Indeed, a quarter of all Tokyo University students come from just eight high schools, six of them private.

Interest extends right down to the 12-year-olds trying to enter elite middle schools. The efforts required for their examinations seem almost unbelievable: successful candidates to top schools average three hours each evening at cramming schools plus 2.5 hours home study. Not surprisingly, sleep often has to be sacrificed: one student studied for seven hours a day after school and slept for only six hours — and the university entrance examinations are six years away.

But time is still found for a prayer. Yushima Tenjin, a famous Shinto shrine close to Tokyo University, has more than 30 metres of high railings from which to hang *ema*, thin boards about 15 cm square on which a prayer can be written. In the examination season, the railings are hidden by tens of thousands of *ema*. "Please help me", "May I succeed", most say. By March, the top layer of *ema* come from those offering thanks for their success. Nearby schoolchildren who have yet to face the examinations can buy lucky Shinto charms to aid study and sets of pencils with inspirational slogans on them. "Effort, effort every day", "Success comes from your own labours".

Later in the year will come the articles that make clear the point of all this effort: those detailing where top ministries and companies recruit their staff: the Ministry of Finance, 89 per cent Tokyo University, the Ministry of Foreign Affairs, 76 per cent Tokyo University... and so it goes on. There is little doubt that the educational system has produced the diligent, untiring staff that government and industry have wanted. The question is whether they will be the right kind of people now that Japan is a world leader in search of originality. That question too is an endless source of magazine articles... **Alun Anderson**

US machine, UK chip

Washington

FLOATING Point Systems (FPS) of Oregon announced earlier this month what it claims to be the world's fastest super-computer, a massively parallel design based on hypercube architecture that employs the "transputer" chip made by Britain's Inmos. Among the first recipients of a Floating Point "T" series machine will be Cornell University, which earlier this year turned down on principle \$10 million of Department of Defense (DoD) funds for a large Floating Point machine because the grant had not been awarded through competitive merit review.

The machine that is going to Cornell is at the small end of the T series range — a 16-node T/20 worth just under \$1 million that can handle 256 million floating point operations per second. The university, which helped FPS develop the machine, has not found a grant to pay for it; a Floating Point spokesman said that if no grant could be found, the two sides would "sit down and talk about it". The top-of-the line T/40000, in contrast, offers a peak speed of 262,000 million floating point operations per second. There are still hopes at Cornell that DoD will soon conclude a merit review for the \$10 million it wants to spend on a large FPS machine and find in Cornell's favour, but some Cornell officials now believe that the university's stand on a point of principle may have cost it the deal; it is rumoured that DoD will put a giant FPS machine at Los Alamos National Laboratory instead and tie it to the University Research Initiative. Los Alamos will not confirm the story, saying only that putting the machine there is "a possibility".

Tim Beardsley

Semiconductor dealing

US and Japan near agreement

Washington

THE US semiconductor industry is fishing for an increase of sales to Japan by diplomatic as well as commercial means. This seems to have been the significance of the flurry of activity last week, during Prime Minister Yasuhiro Nakasone's visit to Washington, when President Reagan himself was cast in the role of part-time chip salesman.

Bilateral talks between the United States and Japan have so far produced no formula for expanding US companies' access to Japanese markets, at least on the scale of Japanese exports of semiconductor chips to the United States. But the mere existence of these talks has raised fears in Europe that the two countries are forming a cartel to carve up the world semiconductor market.

US semiconductor sales are in the doldrums. United Technologies Corporation has closed its Mostek subsidiary, and Intel and Motorola have scaled back their operations. The price of chips has been plummeting. Two years ago, erasable programmable read-only memories (EPROMs) were selling for \$28, but now they cost about \$3 each.

All this has made US companies look more hungrily than ever to a larger piece of the \$8,000–10,000 million Japanese market. While the market share for US semiconductors worldwide is about 52 per cent, sales in Japan have been closer to 10 per cent. Last summer, several US companies asked for and were granted a formal investigation by the US Trade Representative Clayton Yeutter of Japanese trading practices.

Since January, representatives from the Japanese Ministry of Trade and Industry and the Office of the United States Trade Representative held a series of meetings to discuss the US complaints. The Japanese offered to encourage Japanese companies to increase purchases of US chips to around 12.5 per cent, and later increased the target to 19.5 per cent by 1990. But US companies reckon they can grab closer to 30 per cent of the Japanese market under "fair" trading conditions, and the last meeting on 28 March ended with the two sides far apart.

Rumours that a cartel was being planned appear to have stemmed from two factors in the negotiations. Some observers saw the Japanese offer of a specific market share to US companies as a preamble to further international arrangements. This fear was fuelled by the fact that the 19.5 per cent offer was made at a meeting between top executives of Japanese and US companies held separately from the official trade negotiations. But officials of the Semiconductor Industry

Association — a US trade association — dismiss claims of cartel-talk, arguing that they are merely interested in assuring free trade and competition.

Hovering over the trade negotiations is a preliminary decision by the US Department of Commerce that Japan is guilty of dumping chips on the US market. Importers of 64K dynamic random access memory chips (DRAMs) have been posting bonds or paying deposits to cover dumping duties since last December. Those penalties have been expanded to cover EPROMs and 256K DRAMs since last month. If the preliminary dumping charges are upheld, importers of Japanese chips will have to pay as much as 188 per cent of the cost of the chip in penalties. A permanent decision will be issued on 23

April for 64K DRAMs, and on 27 May for EPROMs and 256K DRAMs.

Dumping duties are a double-edged sword. They provide some measure of relief for semiconductor manufacturers, but may alienate their largest customers, the computer and telecommunications industry. Rising prices for semiconductors spell trouble for the entire electronics industry. Moreover, capturing a market by lowering prices is a time-honoured practice in the United States, one that US companies are not shy of following.

The Reagan administration has so far been giving the semiconductor industry strong support in its fight for entry into Japanese markets. Yeutter has warned Japan to expect "harsh actions" if there is no change in trading practices. President Reagan is likely to take a softer tone with Prime Minister Nakasone, but the US industry is hoping his message will be just as clear.

Joseph Palca

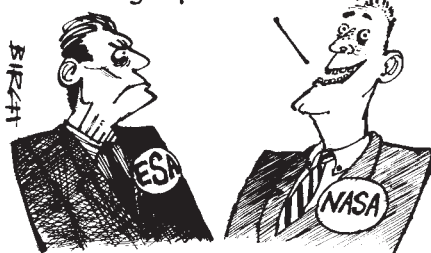
US space station

Europe's centrifugal tendencies

Washington

THE US National Aeronautics and Space Administration (NASA) went to Congress last week to seek \$140 million for design and development of the space station, putting a brave face on its continuing argument with the European Space

Gee, surely Columbus already did enough for the world?



Agency (ESA) on European participation. John Hodge, acting associate administrator for the space station, said he thought agreement with the Europeans in preliminary design work was "very close".

Others are not so optimistic. At issue is whether the pressurized module that ESA will contribute (called Columbus) should be totally dependent on the space station support systems (as NASA wants) or whether it should retain some autonomy that would enable it to be partly free-flying. NASA had wanted the disagreement settled in time for an appearance before the House of Representatives' space station and applications subcommittee, and on 24 March this year Reimar Lüst, director general of ESA, wrote to NASA agreeing that the module would be a permanent part of the space station — although still holding out the hope that it could be detached for research purposes.

The details were still being discussed last week by ESA's member nations, which want the module to have the capability to dock with a free-flying European co-orbital platform and insist that the module be attached to the space station at one end only.

There also seems to be potential for disagreement over the sort of experimental apparatus that will be carried on the European module. Hodge told Congress last week that he expects the European module to be used chiefly for life sciences research, while the US experimental module would concentrate on microgravity research and the Japanese on "technological development". But an ESA representative attending the hearing later said pointedly in private that ESA has every intention of carrying out microgravity research of its own and had not agreed to concentrate exclusively on life sciences.

Canada and Japan have already reached agreement with NASA on their contributions to design and development. Canada will supply a mobile manipulator arm, while Japan is to contribute a pressurized module with an unpressurized experimental area known as "the back porch". NASA's conceptual design for the space station is now significantly less ambitious than was being talked about two years ago; the number of US pressurized modules has decreased from four to two, for example (although their volume has increased). Other changes include the power supply, now planned as 25 kW of solar panels and 50 kW of "solar dynamics" — parabolic Sun-tracking reflectors that focus solar radiation on heat-exchanging liquid that drives turbines.

Tim Beardsley

Deep-sea sonar

New GLORIA in record time

GLORIA III, the latest model of the British deep-sea side-scanning sonar which has fired the imagination of geologists worldwide, has been built in record time at the Institute of Oceanographic Sciences (IOS) at Wormley, Surrey for the US Geological Survey (USGS) and is due to undergo trials off the continental shelf later this week. If all goes well, the converted trawler *Farnella* will set sail with the new GLORIA for the icy waters of the Bering Sea, where GLORIA's sonar beams will be trained onto the sea floor of the United States' northernmost Exclusive Economic Zone (EEZ). Meanwhile, yet another GLORIA is on the drawing boards, but this time to be built by private enterprise.

The new GLORIA's predecessors, I and II, have already mapped more than 2 per cent of the world's ocean floor, including EEZs off the US west coast, Puerto Rico, the Virgin Islands, and in the Gulf of Mexico, revealing many geological features including sinuous deep-sea canyons, hundreds of previously unknown submarine volcanoes, faults, submarine slides, metal-rich deposits and manganese nodule fields. Results from the US west coast were published earlier this month in a huge atlas of glossy black and white sonograph mosaics by USGS using state-of-the-art image-processing techniques.

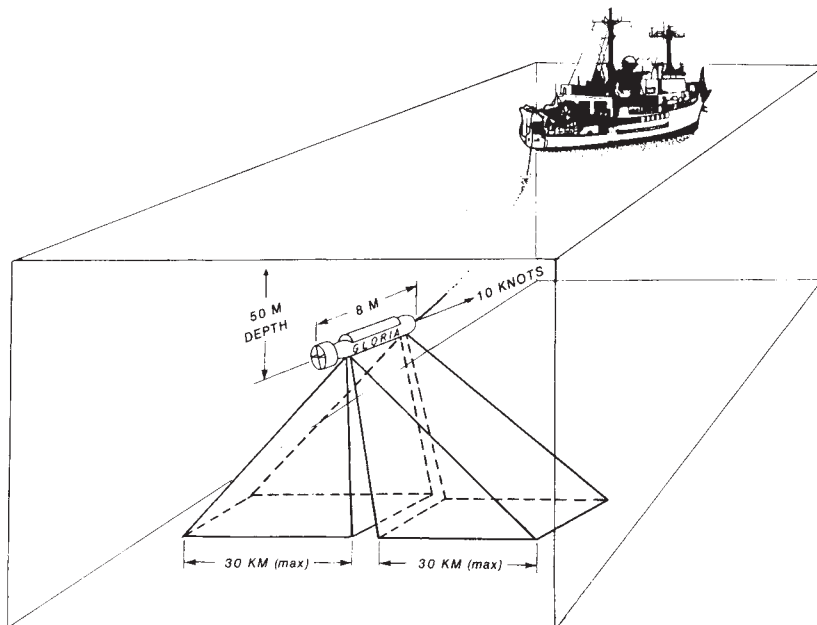
GLORIA, short for Geological Long-Range Inclined ASDIC, is a torpedo-like "towfish" that scans the sea floor with two sonar beams that "illuminate" the sea bed in a swathe up to 30 km wide on either side of the ship track. Two rows of 30 transducers on either side of the towfish emit a 4-second FM "chirp" within the range 6.2 to 6.8 kHz; frequency modulation of the signal allows recognition of the return echo even if the chirp has been half swallowed by the sea floor. The time of return of echoes and their strength are digitally recorded and printed out as sonographs which resemble an oblique aerial photograph but with an important difference. Whereas aerial photographs resolve different objects by registering angular variations, sonar resolves them by measuring differences of slant range. Thus the interpretation of sonographs can be equivocal, but ambiguities are largely resolved by overlapping the swathes to obtain a view of any feature from both sides and by using bathymetric maps to pin down the source of echoes. The sonographs not only reveal local topography but can also indicate the nature of the bottom — sandy beds and manganese nodule fields show up as bright spots because they resound like a bell when hit by GLORIA's sound waves.

Improvements in the Mark III

GLORIA include titanium transducers and a digitized beam-steering unit to correct for yaw, while a new laser printer will be used to produce high-quality sonographs at sea, but these changes are minor compared to the giant stride forward made between Marks I and II. Mark I was a cumbersome 6-tonne Goliath 10m in length and nearly 2m in diameter that had to be manhandled by a team of eight, including divers. Surveys were thus confined to sheltered waters such as the Mediterranean, and GLORIA I could be operated only from RSS *Discovery* which had

the end of September, after which attention will turn to the EEZ around Hawaii. Beyond that, the sea floor around Guam is a future target for insonification, as is the east coast of the United States, but it is unlikely that the two existing GLORIAS (II and III) can cope with the ever-increasing demand for their services, which is where private enterprise may step in.

Marconi have signed an exclusive licensing agreement with the Natural Environment Research Council to produce yet more GLORIAS. Alastair Johnson, director of Marconi's newly-formed Maritime Applied Research Laboratory, sees enormous potential in the Pacific, where several island nations with large EEZs have expressed interest in GLORIA.



GLORIA is neutrally buoyant and is towed about 300m behind the mother ship. On "full beam" at a water depth of 5,000m, GLORIA scans a swath of sea floor about 60 km wide and the time interval between sonar pulses is set at 40 seconds to allow time for the most distant echoes to return; at shallower depths, the area covered and the time between pulses are reduced.

the necessary permanent installations. By contrast, Mark II, a slim featherweight at 2 tonnes and 0.6m diameter, can be deployed from many ships and has successfully been recovered in two hurricanes, including its namesake hurricane Gloria in the Gulf of Mexico. But demand for GLORIA II's services has simply proved too great, and agreement was made to construct a twin for USGS with US funds (see *Nature* 312, 688; 1984). Although the first components for GLORIA III were purchased about a year ago, work on the new model did not get into full swing until December, when the GLORIA team returned from sea, and from late January IOS scientists worked into the night to complete GLORIA on schedule.

After testing GLORIA at "full-beam" off the continental slope, *Farnella* will return to Falmouth on 18 April before heading for San Francisco through the Panama Canal. The survey of the Bering Sea is scheduled to begin in late June and run to

Nations poor in land-based natural resources see a possible bonanza on the sea floor that surrounds them, where metal-rich deposits and manganese nodules are just waiting to be scooped up.

But GLORIA's full potential will probably be realized only when it is combined with other survey systems. For example, the region of dead-space immediately below the GLORIA towfish could conveniently be covered by a higher frequency swathe-mapping system such as SEABEAM or ELAC.

In addition to providing detailed topographic maps and making better use of ship-time, this approach would also help in the interpretation of GLORIA data, by yielding vertical depth (bathymetry) as well as inclined-range data. In its present form, GLORIA is dependent on existing bathymetric maps. Incorporation of a multibeam sonar system would make GLORIA all the more powerful.

David Swinbanks

Dioxin exposure at Monsanto

SIR—Michael Gough¹ says that in our letter² we do not challenge the conclusions that there were no excesses of cancer or heart disease deaths in studies carried out on Monsanto workers exposed to dioxin. This is not true. The fundamental point of our letter is that because of problems in the design of the studies and improper exclusion of some workers, the findings must be treated with some scepticism.

While agreeing with our concern over the indices of dioxin exposure in the two Monsanto studies — because they permit one group to be referred to as dioxin-exposed in one study and as unexposed in the second — Gough suggests that three of the cancer deaths are “likely independent of dioxin exposure”. These were lung cancer cases, all smokers. Although there is no reported association between lung cancer and 2,3,7,8-TCDD exposure, given the cancer-promoting properties of this chemical³, an interaction between TCDD and initiators with respect to lung cancer must be considered⁴. In fact, interactions between TCDD and benzo(a)pyrene (a constituent of cigarette smoke) have been described in animals⁵.

Nineteen workers who died of circulatory disease or cancer were excluded from one epidemiology study by Drs Zack and Gaffey⁶. According to our records, obtained from Monsanto by lawyers for the plaintiffs in a trial against the company, these men should have been included as they were employed by Monsanto between 1955 and 1977. Employment in this period was the criterion for inclusion in the study cohort.

Gough criticizes us for not contacting the authors to find out why these 19 workers were excluded. In fact, we encouraged *Nature* to write to Dr Zack in January 1985 to get her comments on our original letter. She failed to reply to both the January letter and a reminder sent by the journal in February or March.

Gough seems to have misunderstood our point about treating the two Monsanto mortality cohorts — workers exposed to dioxin in the aftermath of the 1945 accident, and hourly paid employees exposed to dioxin during the manufacture of the herbicide 2,4,5-T (ref. 6). According to the 1953 Kettering laboratory study⁸, it is clear that workers in the 2,4,5-T process have the same type of symptoms reported by those exposed in the accident, although the severity of the complaints seems greater in the latter group. Our point is that both groups should be included in epidemiological analysis. Moreover, some of the workers employed on the 2,4,5-T process also helped to clean up after the accident. They had potential exposure to 2,3,7,8-TCDD in both circumstances.

Gough claims that no excess of cancer has been found in the Nitro workers. This is incorrect. A reanalysis of the data, presented by EKS⁹ at the Dioxin 85 Symposium in Bayreuth in September 1985, indicates an excess mortality due to lung and bladder cancer. Similar results on this cohort were also found by an expert retained by Monsanto.

Workers at the Monsanto Nitro plant who were exposed to a bladder carcinogen *p*-aminobiphenyl developed bladder cancer as a result. Some of them may have been exposed to TCDD in 2,4,5-T. Given the cancer-promoting properties of the dioxin, an effect on the incidence and severity of bladder cancer in the Monsanto workers cannot be ruled out either.

We did not suggest that “low-level, intermittent exposures bear more risk than higher-level exposures”, but that health effects such as cancer may be expected to result from lower and more chronic exposure in the absence of chloracne. The report by Moses *et al.*¹⁰ demonstrates that chloracne did not predict work exposure to 2,4,5-T in that nearly 50 per cent of workers with considerable work history with the herbicide never showed chloracne.

On the discrepancy between studies on the use of phenoxy herbicides and chlorophenols in Sweden^{11,12} and New Zealand¹³, it is important to look at the design of each study. The control group in the New Zealand studies consisted of other cancers selected from the National Cancer Registry, which helped reduce recall and interviewer bias, two factors that could have influenced the Swedish findings. In his latest study, Hardell¹² used colon cancer cases as controls to reduce bias which may have occurred in his initial investigation. This study also found an increased incidence of soft-tissue sarcoma¹².

A recent study by Kogan and Clapp¹⁴ showed a higher than expected incidence of soft-tissue tumours in Vietnam veterans, whereas Greenwald *et al.*¹⁵ found no significant association between service in Vietnam and soft-tissue sarcomas. This demonstrates the need to examine each report critically, including the investigations of the Monsanto workers.

As to an excess of deaths due to circulatory disorders in the Monsanto study⁶, the authors calculate proportionate mortality ratios (PMRs) of 137 and 131 for exposed and unexposed cohorts respectively. This excludes 19 additional cases referred to in our original letter². Further, if the authors classified workers as unexposed to 2,4,5-T when they were clearly exposed to dioxin after the reactor accident, this results in misclassification.

Finally, Gough says optimistically that it is “possible to read the literature and

arrive at a cautious conclusion that dioxin has not caused early death and cancer in heavily exposed workers”. We consider this conclusion to be premature.

The purpose of our earlier letter was to put forward reasons why the Nitro data should be reassessed and the cohort followed further.

ALASTAIR HAY

Department of Chemical Pathology,
University of Leeds,
Leeds LS2 9JT, UK

ELLEN SILBERGELD

Environmental Defense Fund,
1616 P Street NW,
Washington, DC 20036, USA

1. Gough, M. *Nature* **318**, 504 (1985).
2. Hay, A. & Silbergeld, E. *Nature* **315**, 102–103 (1985).
3. Pitot, H.C. *Cancer Res.* **40**, 3616–3620 (1980).
4. Weinstein, I.B. in *Public Health Risks of the Dioxins* (ed. Lawrence, W.) 155 (Kaufmann, Los Altos, California).
5. Mattison, D.R. & Nightingale, M.S. in *Public Health Risks of the Dioxins* (ed. Lawrence, W.) 217 (Kaufmann, Los Altos, California).
6. Zack, J.A. & Gaffey, W.R. in *Human and Environmental Risks of Chlorinated Dioxins and Related Compounds* (eds Tucker, R.E., Young, A.L. & Gray, A.P.) 575 (Plenum, New York, 1983).
7. Zack, J.A. & Suskind, R.R. *J. occup. Med.* **22**, 11 (1980).
8. Atkin, R. *et al. Report on a Clinical and Environmental Survey*, Monsanto Chemical Co. Nitro, W.V.A. (Kettering Laboratory, Cincinnati, Ohio, 1953).
9. Silbergeld, E.K. *5th International Conference on Dioxins*, Bayreuth, FRG 16–19 September 1985 (Chemosphere, in the press).
10. Moses, *et al. Am. J. ind. Med.* **5**, 161 (1984).
11. Hardell, L. & Sandstrom, A. *Br. J. Cancer* **39**, 711 (1979).
12. Hardell, L. *Scand. J. Work Environ. Health* **7**, 119 (1981).
13. Smith, A. *et al. J. natn. Cancer Inst.* **73**, 1111 (1984).
14. Kogan, M.D. & Clapp, R.W. *Massachusetts Office of Commissioner of Veterans Services Agent Orange Program*, (Massachusetts Department of Public Health, Division of Health Statistics and Research, 1985).
15. Greenwald, P. *et al. J. natn. Cancer Inst.* **73**, 1107 (1985).

Dowsing explained

SIR—David Marks¹ asserts that there are no theories to account for paranormal effects. This is not true for dowsing. Serious dowsing claims such as those made by Soviet geologists², which are difficult to account for in terms of the reception of normal sensory cues, may be explained by postulating human sensitivity to small magnetic field gradient changes^{3–5}. The theory is supported by a series of tests involving 150 subjects⁶.

The magnetic theory predicts that dowsers can achieve above-chance results only if the features they claim to detect are associated with magnetic gradients of at least one nanotesla per metre. This was not the case in Randi's recent experiments⁷, so his chance results are therefore consistent with the magnetic theory, which merits further investigation.

TOM WILLIAMSON

6 Woburn Court,
Stanmore Road,
Richmond,
Surrey TW9 2DD, UK

1. Marks, D. *Nature* **320**, 119 (1985).
2. Williamson, T. *New Scientist* **81**, 371 (1979).
3. Rocard, Y. *La Recherche* **12**, 792 (1981).
4. Presti, D. & Pettigrew, J. *Nature* **285**, 99 (1980).
5. Baker, R. *Nature* **301**, 78 (1983).
6. Chadwick, D. & Jensen, L. *Utah Water Res. Lab. Prog. Rep.* 78–1 (1971).
7. Randi, J. *Skeptical Inquirer* **8**, 329 (1984).

Brain death in Japan

SIR—Last December, new criteria for the diagnosis in Japan of “brain death” were announced by an advisory panel of the Ministry of Health and Welfare (see *Nature* 318, 591; 1985). I would like to make some further comment and discuss the controversy that has arisen over a transplant operation carried out at Tsukuba. I have been involved in the Patients Rights Conference (PRC) in Tokyo University Hospital for several years. PRC has discussed many cases of medical malpractice and experiments on humans brought to its notice by the victims or their relatives.

The main topic of discussion at present in PRC is brain death. PRC has now held six symposia at Tokyo University to try to understand what brain death really means and those invited have included transplant surgeons, neurosurgeons, neuropsychiatrists, anaesthetists, pediatricians, nurses, lawyers, writers, Diet members, victims of medical malpractice and the handicapped. Recently, we have focused on the case of the first combined pancreas and kidney transplantation (see *Nature* 313, 613; 1985).

According to newspapers and doctors at Tsukuba University, the donor was a neuropsychiatric female patient diagnosed as brain dead after severe brain haemorrhage. The donor had been suffering from psychiatric symptoms derived from cerebrovascular disease and admitted to a local psychiatric hospital. When she went into deep coma, and her husband was told her spontaneous breathing would soon stop, he proposed that doctors should make use of her organs for transplantation, as she had wished.

Immediately after spontaneous respiration stopped, the transplant team began preparations, while the neurosurgeon abandoned treatment before the definite diagnosis of brain death. At first, the team planned to perform a liver transplant but could not find an appropriate recipient. Finally it was decided to perform a combined pancreas and kidney transplant. This quick change raised ethical questions of whether well-informed consent could be obtained from the recipient. The result of the operation was far from satisfactory; the patient died in less than a year from intra-abdominal bleeding from the resected part of the transplanted pancreas. The 29-year-old patient's case (diabetic nephropathy) was presented at the meeting of the Japan Transplant Association last year, at which questions were asked about the adequacy of the diagnosis of brain death, indication of transplant, operation technique and post-operative care. Unfortunately, satisfactory answers were not received from Dr Iwasaki and his colleagues. I should further mention that Dr Iwasaki has written no articles on

pancreas transplantation in his twenty-year career and may have insufficient experience for this procedure.

PRC sees the following problems in the operation: (1) the possibility that necessary treatment for the basic disease of the donor was abandoned; (2) transplant from a brain-dead patient before the establishment of the criteria of brain death; (3) discrimination against the psycho-handicapped whose medical-legal competence was questionable, and the validity of the consent obtained from the guardian; (4) the adequacy of the selection of the recipient who could have survived longer if treated conventionally, with insulin and haemodialysis, for example.

PRC asked Dr Iwasaki for a meeting to discuss the case. Unfortunately he refused, so PRC felt obliged to file the lawsuit against him and his colleagues described in *Nature* in order to have a chance to discuss the matter openly in public.

KATSUNORI HONDA

Faculty of Medicine,
University of Tokyo,
7-3-1, Hongo, Bunkyo-ku,
Tokyo, Japan

Embryo research

SIR—David Davies's letter (*Nature* 320, 208; 1986) makes me more than ever aware that, as the only embryologist on the Warnock Committee of Inquiry into Human Fertilization and Embryology, I was wrong in not insisting on a logical and unambiguous terminology. I missed the first meeting, at which it was decided to apply the term “embryo” to all stages from fertilization onwards, but that is no excuse: plenty of other decisions were changed during the 20 months that the committee sat.

Traditionally, the embryo is the entity that is formed only at the primitive streak stage, the entity which then develops into the fetus and ultimately into the baby. In recent years, embryologists (including me) have adopted the sloppy practice of using the same term for the entire product of the fertilized egg, most of which differentiates before the formation of the primitive streak into tissues that will protect and nourish the future embryo. The ambiguity is well illustrated (unintentionally) in an excellent recent paper describing experimental studies on early marsupial development (Selwood, L., *J. Embryol. exp. Morphol.* 92, 71; 1986). The author refers to “2-cell embryos”, “4-cell embryos”, “14- to 16-cell embryos”, “one embryo, which was a complete blastocyst with about 22 cells”; she then states: “The cells that will later give rise to the embryo [my italics] are at this stage indistinguishable from the rest of the blastocyst”. The embryologist knows what is meant; the uninitiated must be left gasping.

David Davies writes: “... we were all

well aware that the human embryo for the first two weeks of its existence... bears no visual resemblance whatever to the later embryonic and fetal stages”. This misses the point. As defined above, the embryo does not exist for the first two weeks after fertilization. It is formed from a tiny subset of the mass of cells generated during that period by the fertilized egg. Alternative terms for this mass of cells, and any earlier stage back to the fertilized egg, include conceptus, zygote, pre-embryo and pro-embryo (“an embryonic structure preceding true embryo”, *Henderson's Dictionary of Biological Terms*, 9th edn April 1985). All avoid the confusing ambiguity of “embryo”.

“Cosmetic words”, says Dr Davies. Cosmetics hide, clarity illuminates. I strive for clarity, and regret that I have not done so more effectively in the past.

ANNE MCLAREN

MRC Mammalian Development Unit,
Wolfson House, 4 Stephenson Way,
London NW1 2HE, UK

Justifiable risks

SIR—Your leading article “Why did Challenger matter?” (*Nature* 319, 435; 1986) again shows the difficulty of defining “essential”. It says “Last week's tragedy shows... there is now no case for carrying on shuttle flights passengers whose presence is not essential. There should be no more schoolteachers, congressmen or even journalists until experience has shown shuttle flights to be mere routine.”

If our ancestors had taken that view, most of us over here would still be over there in Europe with you rather than here where the opportunities as well as the risks are more abundant!

JAMES A. PITTMAN JR

University of Alabama
School of Medicine,
University Station,
Birmingham, Alabama 35294, USA

More ways than one

SIR—While scientists in Oxford are planning to use genetically engineered viruses to kill the caterpillars of *Panolis flammea* which infect pine trees¹, Chinese scientists in Anhui Province are employing trained grey magpies to eat up pine moths and larvae², a method that seems to be efficient and versatile. Each bird catches about 18,000 pine moths a year and these birds can be moved to another infected area when the moths in one tract are largely exterminated.

PANG-CHUI SHAW

Biotechnology Centre,
Imperial College of Science
and Technology,
London SW7 2AZ, UK

1. Newmark, P. *Nature* 320, 2 (1986).
2. *China Pictorial* No. 2, 42-43 (1986).

Agricultural genetics still on ice

The exciting prospect that biotechnology will transform the technology of agriculture must, sadly, be moderated by the certainty that progress will be unreasonably delayed.

THE accelerating progress of plant biotechnology has been paralleled in the United States by rising disquiet in the research community about disproportionate public concern over the environmental impact of genetically manipulated organisms. The consistent success of Jeremy Rifkin in obstructing field trials of even the most trivially modified organism must have contributed to the anxiety apparent at the Du Pont-UCLA Symposium on Molecular Strategies for Crop Protection two weeks ago, a meeting otherwise notable for its ebullience in the light of recent triumphs (a more detailed account of which will appear in a forthcoming issue).

The meeting, held during the Easter week at Steamboat Springs, Colorado, saw the announcement of two important new developments in the genetic manipulation of crop plants. Hitherto, developments have been largely confined to the successful transfer of marker genes of one sort or another, or of genes conferring tolerance to herbicides. What makes the recent advances potentially important is that they involve the transfer of genes that confer protection from natural pests and pathogens. Biologists at Washington University, St Louis, for example, in collaboration with Monsanto, have used the viral gene encoding the coat protein of tobacco mosaic virus to produce limited protection against the subsequent inoculation of tobacco plants with the complete virus. At the same time, both Agrigenetics and Plant Genetic Systems have demonstrated expression, again in tobacco plants, of the gene for the *Bacillus thuringiensis* toxin, which is lethal to insect pests but harmless to almost everything else.

It is far too early to claim for either of these interesting experiments a place in the history of agricultural practice: although each has come successfully through preliminary tests for efficacy against viruses and insects respectively, greenhouse tests do not reliably predict what will happen in the field; and therein lies the problem. It has taken Agricetus, for example, three years to get permission from the US Department of Agriculture (USDA) and the National Institutes of Health (NIH) simply to test the behaviour in the field of tobacco plants containing nothing more exciting than the vector sequences from the Ti plasmid that is routinely used in plant gene transfer, along with marker genes of no functional significance and an enzyme marker that

makes it possible to keep track of the expression of the transferred genes in the plant. The trials may now take place this summer.

In the meantime, R. Kaufman of Monsanto estimates that it has cost \$2.5 million in risk assessment to develop a soil bacterium, *Pseudomonas fluorescens*, expressing the insecticidal *B. thuringiensis* toxin and intended as a means of biological control of insect pests that attack the roots of plants. Subject to the approval of the Environmental Protection Agency (EPA), the modified organism will make a tentative debut this year, contained within a system of fences and buffer zones that seem somewhat excessive for an organism that is incapable of surviving for more than six months in the soil and can be eliminated by the application of household bleach; especially since the new gene it contains specifies a protein that is already in use as a topical insecticide.

The fear is that such costly delays at each stage in the development of a potential new agricultural product may quite quickly deter even the larger companies from investing further in plant biotechnology. The smaller ones could not possibly survive them.

The arguments in favour of some relaxation of the rules, or at least streamlining of the procedures, range from the frankly sanctimonious (the potential benefit to farmers and mankind in general) to the disarmingly candid (it is in the commercial interests of companies to produce strictly obsolescent bugs), and with an estimated one-third of the world's crops going to feed the world's pests and pathogens it is difficult to question the validity of at least the first of these arguments — assuming, of course, as everyone does, that the genetic manipulation of plants will eventually extend beyond tobacco and its kin.

It would be foolish to argue that no precautions should be taken to limit the spread of genetically modified plants and bacteria, or monitor their possible effects. The problem is that in a sense there is no real problem. By and large, neither the organisms nor the introduced genes are harmful and there is therefore no reason to expect any harm from the combination. What Kaufman and his collaborators have done by way of risk assessment is to establish that the modified *Pseudomonas* is no more toxic than the *B. thuringiensis* that naturally produces the toxin and that the survival of the engineered bacterium is (as

was almost bound to be the case) compromised rather than enhanced by the interference with its genome. What Rifkin and his like seem to be demanding is an assessment of some unknown and certainly unspecifiable risk somehow inherent in the fact of artificially recombining DNA — a risk there is simply no rational way of assessing.

The absurdity of this position is most clearly seen in the notorious case of the ice-minus bacterium whose release by the biotechnology company Advanced Genetic Sciences has been successfully prevented so far by Rifkin and whose future, despite the permit recently granted by EPA, is still in doubt (see *Nature* 319, 254 and 320; 1986).

The point about this particular experiment is that the genetically modified bacterium has a natural counterpart: although the majority of the natural *Pseudomonas syringae* organisms produce a protein that nucleates ice crystals (and can hence damage fruit crops), some lack the gene specifying the ice-nucleating protein.

The same gene is deleted in the genetically manipulated bacteria, and thus the only reason for all the fuss is that in the case of the AGS organism the genetic recombination was deliberately engineered in the laboratory.

History suggests that, with time, the fuss will die down. Anaesthetics for childbirth were a crime against the natural order when they were first used by a few medical pioneers. Indeed in Europe, where for the most part public reactions are more muted, field tests involving the (scrupulously monitored) release of manipulated organisms are scheduled for this year. Britain, for example, will experiment with a genetically modified virus whose release into the pine forests of Scotland is expected to help curb the ravages of the moth *Panolis flammarum* (see *Nature* 320, 2; 1986).

It would be ridiculous to pretend that the protests registered at Steamboat Springs were disinterested; but at the same time it would be a pity if unreasonable restrictions on field testing were to defeat the impetus provided by the interest of industry to fundamental research on important crops such as rice, and important pathogens such as the fungi.

In the end it may be Europe's initiative that saves the day: Americans hate to be left behind.

Miranda Robertson

Astrophysics

Galaxy superclusters and cosmic strings

from Craig Hogan

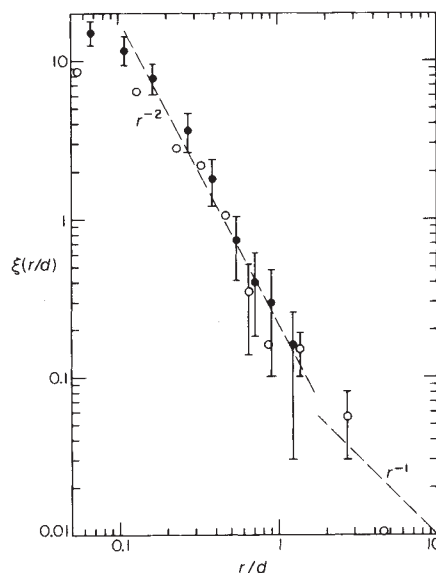
OBSERVATIONAL studies of superclusters — the largest aggregations of galaxies in the Universe — have become increasingly quantitative in recent years, putting a considerable strain on many theoretical models that are unable to account for strong clustering on very large scales. It is thus a considerable achievement for the string theory of galaxy formation that it has now been shown to explain some of the most puzzling observations of superclustering in a remarkably parameter-free way (Turok, *N. Phys. Rev. Lett.* **55**, 1801; 1985).

When the Palomar sky survey was completed in the 1950s, G. Abell examined it by eye and catalogued thousands of the densest aggregations of galaxies, now called Abell clusters. It has since been confirmed that these clusters are themselves clustered on larger scales. The observational controversy about the statistical properties of Abell cluster correlations continues, but little doubt remains about the essential facts: Abell clusters are more strongly clustered than galaxies themselves by a factor of at least ten; and their correlations extend out to much larger scales, again, perhaps a factor of ten in length.

If these correlations are intrinsic (and not the effect of, say, a subtle observational selection effect in Abell's catalogue), the implications are striking. The clusters and the galaxies cannot both be regarded as tracers of a single underlying mass distribution, or their correlations would be the same. Therefore, some factor other than pure gravitational clustering of galaxies must be added to the theory of clustering. It may be possible to explain the correlations in terms of the different types of local conditions and processes that contribute to the creation of a cluster or a galaxy. Thus, galaxies and clusters may represent differently 'biased' samples of the same underlying mass distribution (Kaiser, *N. Astrophys. J.* **284**, L9; 1984). Or, as Turok now proposes, it may be that the phenomenon is a manifestation of something radically different.

As I have previously discussed in these columns (*Nature News and Views* **310**, 365; 1984) cosmic strings are a natural cosmological by-product of certain unified gauge theories corresponding roughly to magnetic flux tubes or superfluid vortices in the cosmic vacuum. Their multifarious unusual gravitational side-effects may include the formation of galaxies (for a review, see Vilenkin, *A. Phys. Rep.* **121**,

263; 1985). Cosmic strings are formed as a dense net of pseudo-random walks, but quickly evolve as the Universe expands into a thin net of infinite (open) strings and a debris of closed oscillating string loops that have broken off from the net. It is the loops which are thought to form galaxies by gravitationally accumulating matter around themselves. Because the



Two-point spatial correlation function ξ for Abell clusters (open circles) and for cosmic string loops (filled circles) as a function of r/d , where r is the separation and d is the mean separation of the objects considered. Error bars in the data for Abell clusters correspond to the square root of the number of pairs in each separation bin (assuming Poisson statistics). Error bars in string simulations correspond to the standard deviations of the mean for 18 independent simulations (from *Phys. Rev. Lett.* **55**, 1802; 1985).

locations of the loops are not determined by gravity, but rather by where and how they broke off from the net of open strings, unusual correlations are expected to be found in the locations of galaxies and clusters. It is these correlations which Turok has now calculated.

Usually correlations among galaxies are thought to be induced by the fact that they attract one another. In the string picture, gravity may play a part in galaxy correlations, but cluster correlations are unlikely to be caused by cluster-cluster attraction but rather by the fact that the string loops causing cluster formation often came from the splitting up of a larger parent loop, and are therefore correlated because of the

(non-gravitational) tension forces within the parent. Such string correlations can be introduced over very large scales because cosmic-string tension is so large that disturbances travel along strings at the speed of light.

To the extent that the statistical properties of the string system evolve self-similarly in time, a scale-free autocorrelation is expected to develop among string loops. Suppose loops of a given size are examined and their correlation length (roughly, the typical size of clusters of the loops) calculated. Then, in a scale-free system, one should find that the correlation length of larger or smaller loops measured at an appropriately scaled epoch should scale as their size, as should their mean separation. This behaviour is in qualitative agreement with what is found — larger, rarer systems such as clusters have larger correlation lengths than smaller, commoner systems such as galaxies. This argument does not establish the ratio of lengths for any given scale or the form of the correlation function, but these details can be calculated by computer. The relativistic equations of motion for strings must be integrated, including the exchange of partners for intercommuting of strings when they cross. By doing this for many initial configurations, it is possible to verify that the network indeed tends towards a state evolving with statistical self-similarity in time, and it is possible to measure the correlations in this state.

A sophisticated computer program has been developed by A. Albrecht and Turok (*Phys. Rev. Lett.* **54**, 1868; 1985) to simulate in this way the evolution of the string network and the formation of loops. Turok then examined the distribution of large loops (large enough to be seeds for the formation of Abell clusters) and calculated their correlation function. Remarkably, his result agrees both in functional form and amplitude with observational results for the Abell cluster autocorrelation (see figure).

It is almost incredible that such a simple prescription (even if the calculational implementation is anything but simple) immediately produces the desired result. It demonstrates the predictive power of the string theory. As calculations become more refined, the theory is likely to confront observation more forcefully. We can expect to see more detailed predictions, such as the correlations of various Abell richness classes, incorporating detailed models of processes such as the gravitational accretion of clusters around loops and the effects of gravitational radiation recoil on loop accretion and correlations. Turok's interesting result provides a framework for such studies. □

Craig Hogan is at the Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA.

Cell biology

Consensus in exocytosis

from Robert H. Kretsinger and Carl E. Creutz

DURING EXOCYTOSIS, the process by which eukaryotic cells release packets of molecules to the extracellular environment, the exocytotic vesicle buds off from the Golgi apparatus and fuses with the plasma membrane, releasing its contents at the time of secretion. Douglas¹ and Rubin² proposed that calcium couples cell stimulation to secretion, analogous to the coupling of excitation to contraction in muscle. Understanding these complex processes has been one of the more rewarding themes of cell biology during the past two decades. On page 636 of this issue³ Geisow and colleagues report the homology of the protein p36 with calelectrin and the closely related proteins pII and endonexin, representatives of a family

ing peptides by chromatography. Several peptides react with antiserum to calelectrin, and Geisow *et al.* sequenced these and several others. Nine different, non-overlapping peptides — one each from p36 and from pII, two from calelectrin and five from endonexin — have similar amino-acid sequences, as summarized in the consensus sequence of the first line of the figure. The carboxy-terminal half of this 17-residue consensus sequence probably forms an α -helix; the first half is predicted to be non-helical.

When preparing this article, we noticed that the sequence of lipocortin (lipomodulin) recently published by Wallner *et al.*⁶ contains three regions homologous with the endonexin fold that is defined by the

Lys-Gly-fob-Gly-Thr-Asp-Glu-var-var-Leu-Ilu-fil-Ilu-Leu-Ala-fil-Arg
 Lipocortin residues 56-72 Met-Val-Lys-Gly-Val-Asp-Glu-Ala-Thr-Ilu-Asp-Ilu-Leu-Thr-Lys-Arg
 Lipocortin residues 128-144 Lys-Gly-Leu-Gly-Thr-Asp-Glu-Asp-Thr-Leu-Ilu-Glu-Ilu-Leu-Ala-Ser-Arg
 Lipocortin residues 287-303 Lys-Gly-Val-Gly-Thr-Arg-His-Lys-Ala-Leu-Ilu-Arg-Met-Val-Ser-Arg

Consensus amino-acid sequences found in endonexin, calelectrin, pII and p36 compared with three homologous domains in lipocortin. Carboxy-termini to the right. Abbreviations: fob, hydrophobic; fil, hydrophilic; var, variable.

that may be involved in the coupling events in secretory cells.

There are many soluble proteins that bind to purified secretory vesicles or to other membranes and components of the cytoskeleton in a calcium-dependent manner *in vitro*^{4,5}. Identification of isoforms and modified forms of these proteins is not complete. Calelectrin, a soluble calcium-binding protein isolated from the electroplax tissue of the ray *Torpedo marmorata* (a tissue very rich in cholinergic synaptic membranes), has a relative molecular mass (M_r) of 34,000. Endonexin ($M_r = 32,500$), another calcium-binding protein, has been identified in the adrenal medulla and in the liver; and the protein pII ($M_r \approx 32,000$) has been identified in the brush border of the intestinal epithelium. But calelectrin, endonexin and pII could be the same protein; antibodies raised to calelectrin have been shown to react with endonexin and, in turn, antibodies to endonexin react with pII. The anti-calelectrin antibodies also react with p70 ($M_r = 70,000$) from liver and with p36 ($M_r \approx 36,000$) from adrenal medulla and brush border.

This cross-reactivity indicates that the five proteins are similar and probably homologous. The new experiments of Geisow *et al.*³ are straightforward and the results confirm these expectations. The authors cut calelectrin, endonexin, pII and p36 into fragments with trypsin or cyanogen bromide and purified the result-

ing peptides by chromatography. Several peptides react with antiserum to calelectrin, and Geisow *et al.* submitted their manuscript but they have noted them in proof. Lipocortin exerts its anti-inflammatory effect by inhibiting phospholipase A_2 , which would otherwise release arachidonic acid from the 2' position of phospholipids. Arachidonate, in turn, is the precursor of leukotrienes and prostaglandins, inflammatory agents.

What is one to make of this amino-acid sequence information? The endonexin fold is not strongly hydrophobic; none of the homologues are inferred to be deeply embedded in a phospholipid bilayer, as are integral membrane proteins. The fold is not homologous with the EF-hand⁷ of calcium regulatory proteins such as calmodulin, troponin C, S-100 and parvalbumin. Lipocortin also appears not to contain an EF-hand. In the EF-hand as well as in extracellular calcium-binding proteins such as thermolysin, the residues coordinating calcium are not adjacent in sequence. The loop of the endonexin fold contains three or four oxygen-containing side chains, depending on which of the nine domains described by Geisow *et al.*³ is considered; however, the Thr-Asp-Glu residues of the consensus sequence are adjacent. One of the domains in lipocortin has only one oxygen ligand, and arginine and histidine are substituted for aspartate and glutamate in this loop. It will be very interesting to know the complete amino-

acid sequence of the endonexin/p36 group of proteins and of synexin, which appears to have activities similar to those of endonexin. If these proteins do bind calcium under physiological conditions, as now seems likely, and if, like lipocortin, they lack EF-hands, they would be the first exceptions to the generalization that calcium-modulated proteins contain EF-hands.

A further interesting development in the calcium modulation story has recently been published. The protein p36, which apparently interacts with chromaffin granules and with components of the cytoskeleton in a calcium-dependent manner, can form a heterotetramer (p36₂, p10₂). The amino-acid sequence of p10 (ref. 8) has two EF-hands. Although in both domains the loops that might bind calcium differ from the archetype, p10 may still impart the calcium sensitivity to the heterotetramer. Perhaps again, as occurred with aqueorin and with calpain, the EF-hand generalization of calcium-binding proteins will be rescued when the complete amino-acid sequences of the other proteins are determined.

But as so often happens, a nice calcium story is sullied by the complication of phosphorylation. One specific tyrosine of p36 (residue 23) is phosphorylated by the tyrosine-specific kinase p60^{src} (encoded by Rous sarcoma virus) and by similar kinases from other transforming viruses⁹. The cellular homologue of p60^{src} is a component of the chromaffin-granule membrane¹⁰ and therefore might phosphorylate p36 during exocytosis. Lipocortin, now known to be a homologue of p36, is also a substrate for a tyrosine-specific kinase. In murine thymocytes, treatment with a mitogen such as concanavalin A, the calcium ionophore A23187 or 4 β -phorbol 12-myristate 13-acetate, leads to a marked increase in the tyrosine phosphorylation of lipocortin. In its phosphorylated form it no longer inhibits phospholipase A_2 and more arachidonic acid is released. Wallner *et al.*⁶ identified Tyr-21 as the probable phosphorylation site of lipocortin; the first endonexin fold begins at Met-56. Will the Tyr-23 of p36 also be 35 residues upstream of its endonexin fold? □

1. Douglas, W.W. *Br. J. Pharmac.* **34**, 451 (1968).
2. Douglas, W.W. & Rubin R.P. *J. Physiol.* **159**, 40 (1961).
3. Geisow, M.J., Fritsche, U., Hexham, J.M., Dash, B. & Johnson, T. *Nature* **320**, 636 (1986).
4. Geisow, M.J. & Burgoyne, R.D. *J. Neurochem.* **38**, 1735 (1982).
5. Creutz, C.E. *et al. J. Biol. Chem.* **258**, 14664 (1983).
6. Wallner, B.P. *et al. Nature* **320**, 77 (1986).
7. Kretsinger, R.H. *CRC Crit. Rev. Biochem.* **8**, 119 (1980).
8. Gerke, V. & Weber, K. *EMBO J.* **4**, 2917 (1985).
9. Glenney, J.R. & Tack, B.F. *Proc. natn. Acad. Sci.* **82**, 7884 (1985).
10. Parsons, S.J. & Creutz, C.E. *Biochem. biophys. Res. Commun.* **134**, 736 (1986).

Robert H. Kretsinger and Carl E. Creutz are at the Departments of Biology and of Pharmacology, University of Virginia, Charlottesville, Virginia 22908, USA.

Planetary evolution

Banded iron formations

from Kurt Lambeck

THE problem of the timescale of the dynamical evolution of the Earth–Moon system just does not go away, as attested by Walker and Zahnle on page 600 of this issue¹. The mathematics of the evolution of a binary planet–satellite system, once it has formed, is reasonably well understood. Much less well understood are the physical mechanisms of dissipation of the energy of the tides raised by the two planets on each other, and this produces very considerable uncertainty in the timescale of any evolution². In the absence of evidence from the distant past, the usual practice is to adopt the modern observed rates of evolution and to extrapolate backwards in time. Despite fluctuating conclusions about the present rate of tidal dissipation, this process leads inexorably to the conclusion that the Moon was very close to the Earth about $2.5\text{--}1.5 \times 10^9$ years ago.

Such a close encounter conjures up planet-disrupting tides in which, to borrow a phrase from the Prayer Book of 1662, “the Earth shall melt away”. Yet the geological evidence for this does not exist. Either the moon escaped unscathed or the close encounter occurred back in the darkest ages of the Earth. It is easier to dismiss the validity of the backwards extrapolation, for there is little reason to suppose that the dissipation rates have remained unchanged with time. Thermal evolution of the solid Earth, evolution of the oceans and the wandering of continents provide sufficient reason to modify the timescale by a factor of two or three.

It is better to follow the chemical arguments. A. E. Ringwood³ has argued for many years that chemical evidence requires that the Moon formed out of the Earth from material evaporated from the mantle during the rapid accretion of the planet, aided by the gravitational energy of the impacting planetesimals that contributed to the construction of the Earth in the first place. Various adaptations of this model are now becoming popular and there is little reason to adopt a starting model for the evolution of the orbital and rotational parameters of the Earth and Moon other than one in which the Moon was close to the Earth in the earliest days of formation. Any information on this evolution then becomes important in understanding the rates of energy dissipation and, by inference, the thermal evolution of the solid planet or the changing volumes and geometries of the oceans.

Different approaches have been used to quantify this evolution². One approach is to model tidal dissipation in past ocean

basins; another is to search the geological record of the Earth for traces of past Earth–Moon configurations. Growth cycles of corals and of molluscs provide some insight into these configurations during Phanerozoic time and the interpretations have received some confirmation from controlled experiments on their living counterparts⁴. Stromatolites, primitive sedimentary organisms from the Precambrian, have also allowed some, less confident, speculation. In both cases, the growth cycles of the organisms are assumed to be controlled by tidal, diurnal and/or seasonal factors during growth and it is also assumed that their record has been preserved intact so that these past cycles can be deciphered.

Now a new observation of past lunar–terrestrial geometries has been proposed. Periodic banding of inorganic material is known to occur in Precambrian time, most notably in the banded ironstones. Walker and Zahnle¹ in their new work suggest that these formations can establish the frequencies of some of the longer cycles in the Earth–Moon system. The banded iron formations of the Hamersley Basin in Western Australia, particularly the finely banded structure of the Weeli Wolli formation, provide the source of Walker and Zahnle’s hypothesis. These bands are primary depositional features, believed to be precipitated from basin waters that were isolated from the oceans⁵. Their regular sequence indicates that their deposition was controlled by periodic natural environmental events. Superimposed on this fine structure are various other band structures, one of which occurs particularly regularly, separated by 23.3 (on average) of the finer bands⁵. Trendall^{6,7} favours an annual frequency for the micro-bands, with deposition controlled by parameters such as water temperature, possibly with a biological involvement. This deposition sequence is then modulated or interrupted by processes of longer periodicity.

Trendall notes that the proposed periodicity of the mesoscale cycle is close to that of the double-sunspot or Hale cycle, although he points out that the absence of an 11-year cycle places this interpretation in some doubt. Note, however, that several indicators of modern climatic trends exhibit pronounced double-sunspot cycles and only weak 11-year cycles⁷.

Other evidence for apparent solar periods in the geological record has recently been presented by Williams and Sonett in the form of sedimentary cycles in late Precambrian varves of South Australia⁸, although here both the 11- and 22-year

cycles were identified. The authors suggest their observations constrain models of the physics of the solar activity cycle.

Walker and Zahnle¹ offer a different interpretation. They suggest that the mesoscale bands are controlled by the lunar nodal cycle, which at present has an 18.6-year period. If correct, the banded ironstones, dated at about 2.5×10^9 years, provide a valuable constraint on the orbital evolution of the Moon. The authors conclude that a nodal period of 23.3 years at the time of the deposition requires that the Moon was at about 50 Earth-radii from the Earth, compared with the present distance of about 60 Earth-radii. Protagonists of early formation of the Moon who have been disturbed by the timescale problem can now rest in peace.

But the basic assumption made by Walker and Zahnle may provoke argument — an essential question is whether the nodal cycle is sufficiently significant to leave its mark on climate. Walker and Zahnle quote several analyses that purport to show that it does. But a solar influence may be more reasonable. For example, is there a similarity between the cycle of these mesoscale bands and the 23-year cycle found in the recurrence of prolonged droughts in North America⁹? Can severe evaporate concentration cause precipitation of the silica-rich mesoscale bands or can a change in wind patterns set up a new circulation system in the water body?

Walker and Zahnle do propose a test of their hypothesis. They suggest that some of the Milankovitch periodicities of orbital motions (periodicities that have a prominent role in modulating the ice age climate) will also have been different in the past from in the present and that this difference may be reflected in periodicities of the mesoscale bands. A macro-band structure has been noted in the banded iron formation, as well as a cyclicity of mesobands intermediate in scale between the macro- and meso-banding^{5,6}. These cycles may offer a test of the new hypothesis, but it is unlikely that sufficient records exist. Alternatively, such a test may reveal some of the longer cycles such as the approximately 100- or 200-year cycles seen in modern climatic records⁷. Either way, further examination of the Weeli Wolli and similar formations is called for. □

1. Walker, J.C.G. & Zahnle, K.J. *Nature* **320**, 600 (1986).
2. Lambeck, K. *The Earth's Variable Rotation* (Cambridge University Press, 1980).
3. Ringwood, A.E. *Origin of the Earth and Moon* (Springer, Berlin, 1979).
4. Scrutton, C.T. in *Tidal Friction and the Earth's Rotation* (eds Bosch, P. & Sundermann, J.) 154 (Springer, Berlin, 1978).
5. Trendall, A.F. in *Iron Formations: Facts and Problems* (eds Trendall, A.F. & Morris, R.C.) 69 (Elsevier, Amsterdam, 1983).
6. Trendall, A.F. *Econ. Geol.* **68**, 1089 (1973).
7. Lamb, H.H. *Climate: Present Past and Future* (Methuen, London, 1972).
8. Williams, G.E. & Sonett, C.P. *Nature* **318**, 523 (1985).

Kurt Lambeck is Professor of Geophysics at the Research School of Earth Sciences, Australian National University, Canberra 2600, Australia.

Fossil record

Magnetic skeletons in Davy Jones' locker

from Richard B. Frankel

MANY fossils of ancient plants and animals are embedded in sedimentary deposits but, by comparison, the fossil record for bacteria is sparse. Now, however, Petersen *et al.* on page 611 of this issue¹ and Stolz *et al.*² present evidence that magnetotactic bacterial microfossils litter the deep ocean sediments.

Various species of magnetotactic bacteria inhabit freshwater and marine sediments throughout the world³. Magnetotaxis (orientation and migration along geomagnetic field lines) is based on magnetosomes, chains of membrane-enclosed magnetite (Fe_3O_4) particles in cells. The chain of magnetosomes imparts a permanent magnetic dipole moment of the order of 10^{-12} electromagnetic units to each cell, sufficient for it to be passively oriented in the geomagnetic field at ambient temperatures. Thus magnetotactic bacteria are self-propelled magnetic compasses — as E.M. Purcell once remarked, “this is a case of the compass steering the ship.”

Magnetosomes and their arrangement in the cell are a masterpiece of permanent magnet design⁴. First, magnetosomes have a narrow size distribution, typically 400–1,000 Å, which is in the permanent, single magnetic size range for Fe_3O_4 . They are too small to lower their magnetostatic energy by forming magnetic domains, yet are too large to be superparamagnetic, that is, for thermally activated relaxation of their magnetic dipole moments. Second, dipole-dipole interactions between the magnetosomes in a chain cause the individual particle moments to orient parallel to each other along the chain direction. Thus the whole magnetosome chain is an elongated, permanent, single magnetic domain.

The single magnetic domain properties of the magnetosome chains can persist in sediments as magnetic microfossils after the bacteria die, making a substantial contribution to the remanent magnetization of the sediments. In their new study, Petersen *et al.*¹ report fossil magnetosome chains in deep-ocean sediments from the South Atlantic dating from 5 to 50 million years ago. The identification is primarily based on the narrow size distributions, shapes and arrangements of the magnetite particles extracted from the sediments. Magnetotactic bacteria are known to produce magnetosomes with diverse, species-specific morphologies^{3,5,6} which can be distinguished from nonbiogenic magnetite. In a related paper, Stolz *et al.*² report living magnetotactic bacteria and magne-

some chains in the upper layers of deep-sea sediments from the Santa Barbara basin. Blakemore⁷ and Kirschvink and Chang⁸ previously reported magnetosome-shaped particles in sediments, but Petersen *et al.* and Stolz *et al.* now claim that magnetosomes are the primary carrier of stable magnetic remanence in deep-sea sediments.

How far back in the geological record might one expect to find magnetic microfossils? Magnetotactic bacteria are microaerophiles, requiring trace amounts of molecular oxygen for magnetosome formation⁹. Thus they currently inhabit the relatively restricted microaerobic zone in sediments. Blakemore *et al.*⁹ argue that originally magnetosome formation would

have been possible only in microhabitats such as algal mats where O_2 was being produced. However, as free O_2 began to accumulate in the global atmosphere about 2×10^9 years ago, the Earth's surface would have been microaerobic by today's standards and microaerophiles, including magnetotactic bacteria, might have been the most prevalent organisms. Thus they may have played a major part in the deposition of iron-rich minerals including magnetite in the widespread banded iron formations that accumulated at that time. □

1. Petersen, N. Dobeneck, T.V. & Vali, H. *Nature* **320**, 611 (1986).
2. Stolz, J.F. Chang, S.B.R. & Kirschvink, J.L. *Nature* (in the press).
3. Blakemore, R.P. *A. Rev. Microbiol.* **36**, 217 (1982).
4. Frankel, R.B. *A. Rev. Biophys. Bioengng.* **13**, 85 (1984).
5. Matsuda, T. *et al.* *Nature* **302**, 411 (1983).
6. Mann, S. *et al.* *Nature* **310**, 405 (1984).
7. Blakemore, R.P. *Science* **190**, 377 (1975).
8. Kirschvink, J.L. & Chang, S.B.R. *Geology* **12**, 559 (1983).
9. Blakemore, R.P. *et al.* *Geomicrobiol. J.* **4**, 53 (1985).

Richard B. Frankel is a Senior Research Scientist at the Francis Bitter National Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Inbreeding

Do animals avoid incest?

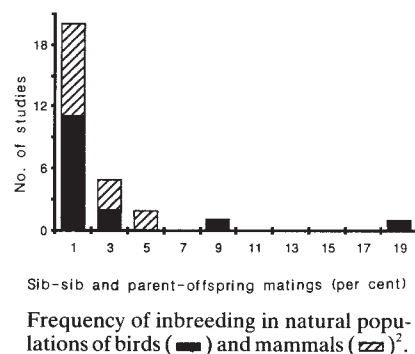
from Paul H. Harvey and Katherine Ralls

SEVEN years ago in these columns, Robert M. May wrote a prophetic article entitled “When to be incestuous”¹. He reviewed the then current literature on inbreeding depression and inbreeding avoidance in human and other animal populations. He went on to develop a model that estimated the likely costs of mating with relatives of differing degree. But the connections between model and data were tenuous, and May argued that “piece after piece can too

breeding avoidance as a force influencing mating patterns in natural populations of vertebrates. Various new studies provide the data to answer some of these questions.

How common is incest? If we define incest as matings between parents and their offspring or between siblings, then a review of many long-term field studies of birds and mammals reveals that incestuous matings are usually less than 2 per cent of the total² (see figure). This compilation refutes one claim³ that “Relatively high frequencies (10–100 per cent of matings) of parent–offspring or full sibling incest have... been documented in natural populations of... various vertebrates”. The six studies of birds and mammals that were cited in support of the statement either show a frequency of less than 10 per cent or contain no relevant data at all. The only reported study with a frequency of incestuous matings in the 10–100 per cent range is of 19.4 per cent in a population of the cooperatively breeding splendid wren (*Malurus splendens*)².

Knowledge of the frequency of incestuous matings is of limited use because we do not know the frequency that would be expected if such matings are not avoided. For example, in most species of birds and mammals, males and females move different distances from their birthplace to their breeding site⁴. Consequently, matings between siblings become less likely. Are



easily be added to build an enchanting castle that rises free, constrained to earth with too few anchorlines of fact”. He concluded with a plea for more data on kinship structure and other relevant variables from natural populations. Simply because of the lack of data, several unresolved arguments have since appeared in the literature about the importance of in-

such sex differences in dispersal to be used as given when formulating the statistical null hypothesis that mating is at random? To answer that question we need to know whether sex differences in dispersal are maintained in populations as incest avoidance mechanisms, or if they may be explained in some other way. There is considerable disagreement over this issue.

Moore and Ali⁵ are particularly vocal critics of the inbreeding avoidance hypothesis. They argue instead that other factors, such as intrasexual competition, more readily explain the available data. Their evidence is challenged by Packer⁶ who cites several results that do not appear to be adequately explained by Moore and Ali's ideas, but which do accord with the inbreeding avoidance interpretation of dispersal. For example, in several studies of primates, males actually receive more harassment from the group they join than they did from the one they left. If, as Moore and Ali argue, males emigrate as a result of aggression from other members of their natal troop we should not expect them to join new troops where conditions are even worse. Nevertheless, Packer concludes, as did Moore and Ali before him, that it may be unwise to prejudge the issue as there are so few data available.

But behavioural observations of incest avoidance are providing increasing support for the inbreeding avoidance hypothesis. Do sexually receptive adults avoid incestuous matings if the opportunity arises? Good data are still rare, but studies on acorn woodpeckers (*Melanerpes formicivorus*)⁷, prairie dogs (*Cynomys ludovicianus*)⁸ and chimpanzees (*Pan troglodytes*)⁹ all provide clear evidence that animals avoid incestuous matings. For example, in Hoogland's study⁸, female black-tailed prairie dogs that matured in groups when male relatives were the dominant breeding males used various methods to avoid mating with them — mating with unrelated males in the group; leaving the group to mate elsewhere; failing to copulate while in oestrus; or even failing to come into oestrus.

One recent report uses the astonishing life-history of the mouse-like marsupial genus *Antechinus* to provide further support in favour of the inbreeding avoidance hypothesis for male-biased dispersal in mammals¹⁰. In both *A. swainsonii* and *A. stuartii*, all one-year-old males die a few days after the mating season. When the females give birth, there are no adult males to be found. Nevertheless, all young males disperse to mate. They are not forced to disperse by dominant males, nor do they leave male siblings behind to sire offspring in the group. The only adaptationist interpretation left seems to be inbreeding avoidance. In addition to sex differences in dispersal, the male die-off in *Antechinus* has provided a novel life-history pattern which can be used to test

other evolutionary ideas about ageing¹¹, litter size and sex-ratio variation¹².

Nevertheless, in those species where sex differences in dispersal exist, the chances of incestuous matings could remain low even if the mate of an animal is chosen irrespective of genetic relatedness from those available near the chosen breeding site. Data from a Dutch population of the great tit (*Parus major*) are in accord with this hypothesis¹³. Females disperse longer distances than males, but mates are related to each other on average by the same degree that males are related to females in adjacent territories. At least one British population of the same species may differ in this respect. Juveniles tend to learn song types typical of those sung at their birthsite, and preliminary data indicate that animals with similar song avoid mating with each other¹⁴.

The extent of inbreeding depression in natural populations is still virtually unknown, but it is becoming increasingly

clear that birds and mammals do not generally mate with close relatives. □

1. May, R.M. *Nature News and Views* **279**, 192 (1979).
2. Ralls, K., Harvey, P.H. & Lyles, A.M. in *Conservation Biology: Science of Diversity* (ed. Soule, M.) (Michigan University Press, in the press).
3. Shields, W.M. *Philopatry, Inbreeding, and the Evolution of Sex* (State University of New York, Albany, 1982).
4. Greenwood, P.J. *Anim. Behav.* **28**, 1140 (1980).
5. Moore, J. & Ali, R. *Anim. Behav.* **32**, 94 (1984).
6. Packer, C. *Anim. Behav.* **33**, 666 (1985).
7. Koenig, W.D., Mumme, R.L. & Pitelka, F.A. *Z. Tierpsychol.* **65**, 289 (1984).
8. Hoogland, J.L. *Science* **215**, 1639 (1982).
9. Pusey, A.E. *Anim. Behav.* **28**, 543 (1980).
10. Cockburn, A., Scott, M.P. & Scotts, D. *Anim. Behav.* **33**, 908 (1985).
11. Diamond, J.M. *Nature News and Views* **298**, 115 (1982).
12. Lee, A.K. & Cockburn, A. *Evolutionary Ecology of Mammals* (Cambridge University Press, 1985).
13. van Noordwijk, A.J., van Tienderen, P.H. & de Jong, G. *Naturwissenschaften* **72S**, 104 (1985).
14. McGregor, P.K. & Krebs, J.R. *Nature* **297**, 60 (1982).

Paul H. Harvey is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK; Katherine Ralls is in the Department of Zoological Research, National Zoological Park, Smithsonian Institution, Washington DC 20008, USA.

Neuroanatomy

Problems in pineal physiology

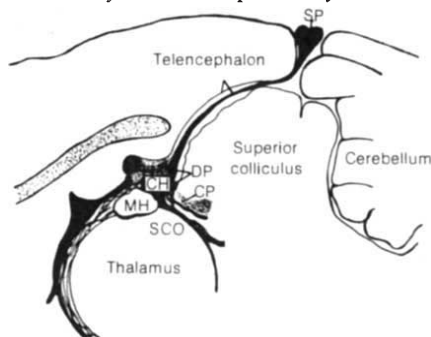
from Michael Hastings

THE pineal of lower vertebrates functions as an endocrine organ, a biological clock and the classical 'third eye'. But in mammals its only identified function is that of a sensory transducer in which melatonin is released into the circulation as a result of circadian information. In photoperiodic species, this signal causes changes in hormonal and metabolic activities that adapt the animal to a seasonal environment (*Photoperiodism, Melatonin and the Pineal*; Ciba Fdn Symp. Vol. 117). The accepted view is that communication between the pineal and the brain occurs by this exclusively hormonal output. However, H.-W. Korf *et al.* (*Science* **231** 717; 1986) now report that immunocytochemically identified pinealocytes of the

Djungarian hamster (*Phodopus sungorus*) send processes similar in structure to nerve axons to adjacent areas of the brain. This result may require a reappraisal of our views on the phylogenetic developments of the pineal and its physiological role in mammals.

In lower vertebrates, the presence of photoreceptor cells, sensory neurones and a well-defined sensory nerve justify the description of the pineal as a third eye. Image formation is poor, but the gland is essential for rhythmic adaptations to the light-dark cycle. In some species, this has led to the development of an endogenous circadian clock within the pineal. In the course of evolution, the mammalian pineal has retained a functional association with the light-dark cycle, shown by the pronounced circadian rhythm of melatonin production, but it has no intrinsic oscillatory properties and neural control of its melatonin rhythm is mediated by a sympathetic input. It is not clear why the autonomic nervous system is involved in brain-pineal communication.

The embryonic development of the mammalian pineal hints at the evolutionary past of the gland. Fetal mammals, including humans, develop a short-lived nerve tract equivalent to the pineal sensory nerve of poikilotherms. Pinealocytes of the fetal rat express a transient period of photosensitivity and those of the adult display spontaneous electrical activity and have ultrastructural features, such as synaptic ribbons, similar to those of photoreceptor cells. In a few species,



Simplified diagram of a sagittal section through the brain of the Djungarian hamster. SP, superficial pineal organ; PS, pineal stalk; DP, deep pineal organ; CP, posterior commissure; CH, habenular commissure; MH, medial habenular nucleus; SCO, subcommisural organ. (Redrawn from *Science* **231**, 737; 1986.)

there are neurones in the pineal parenchyma of adults, located either singly or in a ganglion. Although it is unclear whether they are of pineal origin or are derived from the habenula, it is now thought that these neurones and adjacent secretory pinealocytes are in synaptic contact with central fibres innervating the gland via the habenula and posterior commissures (see figure) (Korf H.-W. & Moller, M. *Pineal Res. Rev.* 2, 41; 1984). This necessitates modification of the established model of solely autonomic regulation of pineal activity. Adding another twist to the problem of pineal-brain communication, Korf *et al.* in their new work use an antiserum against bovine retinal S-antigen (a specific marker for photoreceptive elements) to stain cells in the pineal of the Djungarian hamster. Expression of the antigen by presumed secretory pinealocytes supports the earlier conclusion, based on comparative anatomical and developmental evidence, that these cells derive from an ancestral photoreceptor. Immunoreactive cells extend throughout the pineal but scattered elements are found in the habenula. These scattered cells may be pinealocytes in an aberrant position, but the possibility that some habenula neurones express the antigen cannot be excluded. How can this problem be resolved?

The pineal of the Djungarian hamster is divided into deep and superficial parts connected by a narrow stalk (see figure). The deep pineal and habenula share the autonomic innervation of the superficial gland, but the deep pineal has no rhythm of melatonin production and is not involved in photoperiodic responses. Korf *et al.* observe that the labelled cells of the deep pineal project to the adjacent habenula and posterior commissure. The varicose morphology of the projections is similar to the form of neuronal axons commonly encountered elsewhere in the central nervous system, which implies that the projections form an anatomical basis for neural communication between electrically active secretory pinealocytes and the brain. This novel form of output appears to strengthen the functional association of the mammalian pineal with that of the poikilotherms.

But, if these processes do represent a neural output, such afferent innervation is very different from that of lower vertebrates and fetal mammals. First, the pinealofugal processes are extensions of the pinealocyte, and not of ganglion cells; second, the projections do not form a single, definable nerve tract but seem to run irregularly throughout the epithalamic region; and finally, there is no evidence for sensory receptor function in the mammalian pineal. Consequently, information relayed along this route would presumably be interoceptive and one can only speculate on its function. If pinealocytes

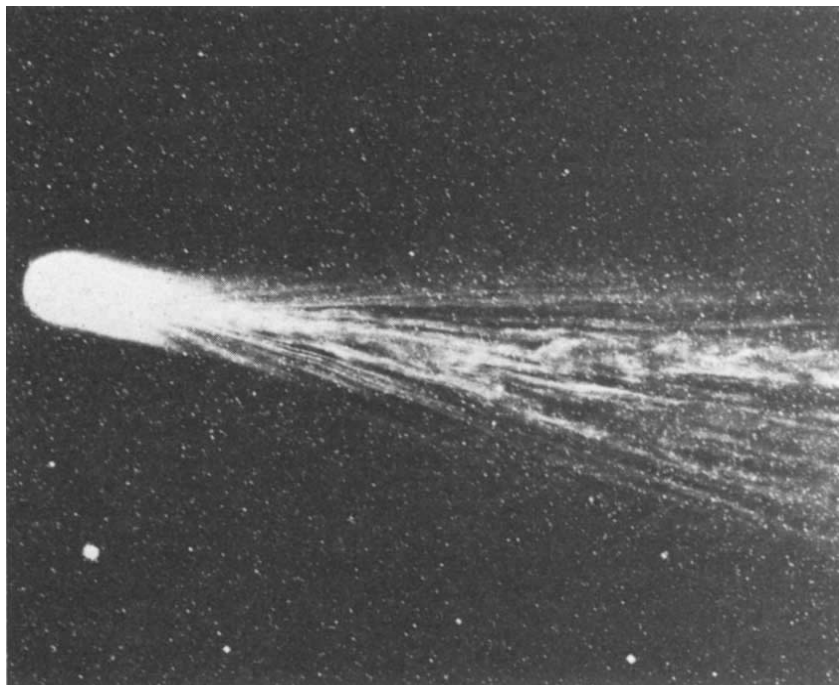
are scattered throughout the epithalamus, is it possible that the connections are responsible for the synchronization of metabolic activity? Alternatively, the processes may indeed be neurally active, conveying afferent impulses to synaptic contacts with central neurones.

Korf *et al.* speculate that melatonin is a transmitter in the centrally directed processes. But pinealocytes are known to contain large amounts of unidentified secretory material, particularly peptides, and there is no *a priori* reason why melatonin should be the only product that can fulfill this function. Indeed, the biosynthesis and release of melatonin from the superficial pineal suggests that specialized storage and release mechanisms, a characteristic feature of conventional neurotransmitter systems, do not exist in this case. Melatonin released from the processes may have a paracrine function in the region of the epithalamus parallel to its proposed control of pigment migration in photoreceptive structures. Such an effect in the brain would be unusual and available data suggest that melatonin acts in the hypothalamus, some distance away.

To make functional sense of the putative central projections of the pineal that I have described here, there must be a precise description of their sites of termination and synaptic associations with central neurones. In addition, to understand pineal physiology in mammals, functions other than photoperiodic time measurement and the melatonin signal must be identified. Should pineal peptides other than melatonin be shown to have regulatory functions, these transmitters could mediate a central action complementary to the humoral effect, following the precedent of the small neuropeptides of the hypothalamus. Of course, parallels between the epithalamus and hypothalamus were suggested years ago (Roussy & Mosinger *C.r. Soc. Biol., Paris* 127, 655; 1938). Perhaps these early anatomists were able to use perceptive pathways no longer exercised in their modern counterparts. □

Michael Hastings is a demonstrator in the Department of Anatomy, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK.

Structures in the gas tail of comet Halley



The delicate and complex details in the ion tail of comet Halley are revealed in this photograph taken on the 1.2-m UK Schmidt Telescope at Coonabarabran, New South Wales, Australia. The original 14-inch square plate has been subjected to special image enhancement techniques by David Malin of the Anglo-Australian Observatory to reveal this amount of detail which is normally hidden in the bright shroud of the dust tail. The structures seen here are produced by the interplay of the Sun's radiation and magnetic field on molecules released from the tiny cometary nucleus. The comet was at a distance of about 150 million kilometres when this 15-minute exposure was made. The bright fuzzy object near the head of the comet is the globular cluster Messier 75 at a distance of about 100,000 light years. First results from the Soviet, Japanese and European spacecraft encounters with comet Halley will be published in *Nature* on 15 May.

Quantum mechanics

Bulletins delivered from the quantum world

from Abner Shimony

MORE physicists at present than in the past seem to be discontented with using quantum mechanics merely as a tool for ordering phenomena. There is a widespread desire among those who apply quantum mechanics to understand it fully to their own satisfaction and many investigations have been stimulated by the conceptual difficulties of the theory. There have been attempts to devise new interpretations or alternative theories, such as theories of hidden variables, which would eliminate these difficulties, as well as attempts to formulate a philosophical world-view that would accommodate them in a natural manner. Many experiments have been performed for the purpose of testing quantum mechanics against alternative proposals, or just to see whether quantum effects can be detected in new domains. Investigations of all of these types were represented at a recent meeting* in New York, which was the first major international conference on the foundations of quantum mechanics to take place in the United States.

Superposition

Most of the conceptual difficulties of quantum mechanics result from the superposition principle, which permits a physical system to have states in which a quantity (such as position or energy) does not have a definite value. When a measurement of that quantity is performed a definite value is found, but it is obscure how the mathematical laws of quantum mechanics permit a definite experimental result. Quantum mechanics also predicts that experiments performed on a properly prepared pair of spatially separated particles can yield results which are correlated in a way that is inexplicable by classical physics—a characteristic known as non-locality. At the conclusion of the meeting there was a widespread sentiment (although not a consensus) that these difficulties are not fully understood. Nevertheless, the theory which generates these difficulties remains well entrenched, for the most beautiful experiments reported fully agree with quantum mechanics.

There were many reports of experimental results in optics, neutron interferometry, electron interferometry and magnetic phenomena. One such experiment was the coherent splitting of a beam

of neutrons polarized in the z -direction into two beams polarized respectively in the x and $-x$ directions and then recombined to exhibit the initial z -polarization (H. Rauch, Atominstitut, Vienna). Also described were an interference phase shift in a neutron interferometer, which agrees with quantum-mechanical calculations of the Sagnac effect—the effect of the Earth's rotation (S. Werner, University of Missouri); and the realization of a 'delayed choice' experiment in an optical interferometer (C. Alley, University of Maryland; A. Zajonc, Amherst College). In this experiment, a Pockels cell switch, activated after a single photon has been split in an interferometer, decides whether there will be an observation of which path the photon is in or of a coherent combination from both paths. These and other experiments described at the meeting agree with quantum mechanics in a good, sometimes spectacular, manner.

But several experiments in progress are motivated by doubts about the decisiveness of previous experiments which have been widely accepted as excluding local hidden-variables theories. A notable example is the experiment of A. Aspect and collaborators, which tests Bell's inequality in an apparatus with polarizer setting chosen while the two photons are in flight. J.-P. Vigiér (Institut H. Poincaré) predicted that a revised version of Aspect's experiment, to be performed at Catania, would show non-quantum-mechanical correlations when the time interval between the detections of the photons is sufficiently small.

J.D. Franson (Johns Hopkins University) has constructed a 45-metre optical interferometer, for which he calculates that a hidden-variables theory would permit interference results in agreement with quantum mechanics only at the price of a violation of relativity theory. A completed experiment, using a shorter interferometer, does agree with quantum mechanics (P. Grangier, Institut d'Optique, Orsay), but Franson claimed that this experiment fails to exclude a local hidden-variables alternative to quantum mechanics, both because of the shortness of the interferometer and because of the failure to use a single-photon light source, as he is planning to do.

There was much discussion about the possibility of reconciling quantum-mechanical non-locality with relativity theory if the earlier non-local results are sustained by future experimentation. It was gener-

ally agreed that quantum-mechanical non-locality has the peculiarity that a signal cannot be transmitted faster than light, and in this sense there is 'peaceful coexistence' with relativistic space-time structure; but there is no consensus that a full understanding of this coexistence has been achieved. Furthermore, there are great difficulties in achieving a special relativistic quantum-mechanical treatment of the measuring process, because an actual measurement occupies a space-time region, which can have arbitrarily great spatial or temporal extent, according to the frame of reference (E.P. Wigner, Princeton University). Wigner also noted the profound difficulties that quantum mechanics poses for general relativity, which postulates that space-time structure is physically variable. These considerations are part of his general contention that the domain of validity of quantum mechanics must be limited.

Limitations

If quantum mechanics does indeed have limited validity, one reasonable place to look for its breakdown is at the level of macroscopic bodies, which are normally amenable to description by classical physics. A. Leggett (University of Illinois) and several speakers associated with him have explored the feasibility of observing macroscopic quantum-mechanical effects (see Eckern, *U. Nature News and Views* 319, 726; 1986 for a discussion of this point).

The magnetic flux trapped in a superconducting ring is macroscopic in the sense that it can be measured directly by a magnetometer without amplification devices, and yet it may exhibit quantum-mechanical behaviour in two ways: tunnelling of trapped flux through a potential barrier; and coherent superposition of two distinct flux states (by analogy to the superposition of chiral states in ammonia). Experimental evidence for the occurrence of genuine quantum-mechanical tunnelling, distinguishable by temperature dependence from classical thermal barrier penetration, has been obtained (S. Washburn, IBM Watson Research Center).

If quantum mechanics is confirmed at the macroscopic level, which most of the participants seemed to think probable, then the great problems of interpreting the theory will in one way become more intractable and in another less pressing for working physicists—according to the old Viennese maxim, "the situation is hopeless but not serious". But there is little doubt, in view of the intensity of interest exhibited at the conference, that imaginative and responsible speculation on these problems will continue. □

Abner Shimony is in the Department of Physics and Philosophy at Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA.

*Techniques and Ideas in Quantum Measurement Theory, 21–24 January 1986 at the World Trade Center, New York. Sponsored by the New York Academy of Sciences and dedicated to Professor E.P. Wigner.

Catabolic plasmids, not dissimilatory

SIR—A recent letter¹ proposed that the term “dissimilatory plasmids” be used to describe the bacterial plasmids that encode pathways for the conversion of compounds such as toluene, naphthalene and camphor to central metabolites. Whilst agreeing with the authors that a single, generally accepted term for such plasmids should be used, I disagree with their solution of abandoning all three of the current terms in favour of a fourth.

Certainly ‘degradative’ does not necessarily imply the complete breakdown of the growth substrate and ‘metabolic’ could imply biosynthetic functions. However, it is my understanding that ‘catabolic’ exactly covers the function of these plasmids, namely either the partial² or the complete³ breakdown of the growth substrates for the purposes of fuelling central metabolism. ‘Dissimilatory’ does not appear to supply any additional nuances of meaning, and suffers the double disadvantage of being difficult to enunciate from the lecture rostrum and subject to incorrect spelling, even by learned journals⁴.

PETER A. WILLIAMS

Department of Biochemistry & Soil Science,
University College of North Wales,
Bangor, Gwynedd LL57 2UW, UK

1. Balajee, S., Boominathan, K. & S. Mahadevan, A. *Nature* **719**, 728 (1986).
2. Fennelwald, M., Benson, S., Oppici, M. & Shapiro, J. J. *Bact.* **139**, 940 (1979).
3. Worsey, M.J. & Williams, P.A. *J. Bact.* **124**, 7-13 (1975).
4. Wheelis, M.L. *A. Rev. Microbiol.* **29**, 505-524 (1975).

Relationship of *Blym* genes to repeated sequences

SIR—The *Blym-1* gene of chicken was defined as a fragment of chicken lymphoma DNA which produced oncogenic transformation in mouse fibroblast cells¹. From the DNA of mouse cells thus transformed, a clone of *Blym-1* was isolated, identified by its transforming activity, and sequenced². However, I now report that the sequence which was presented is actually a fragment from a very abundant dispersed repeat family in mouse DNA.

The repeat family is LINE-1 (ref. 3). Homology to “chicken *Blym-1*” was first discovered⁴ in a computer search with the distantly related human LINE-1 sequence, but is much more striking with the mouse LINE-1 sequence. The whole 590 nucleotides of “chicken *Blym-1*” aligns with nucleotides 2,402–2,995 of the mouse sequence of ref. 5, with only 10.5% mismatch (including 11 insertions or deletions of single nucleotides and one of four nucleotides). Over major parts of the sequence the mismatch is only 7.5%. These values are typical for divergence

between dispersed repeats within a single species. Indeed “chicken *Blym-1*” gave a very strong dispersed hybridization to mouse DNA on Southern blots². Although some dispersed hybridization was also seen to chicken DNA², its origin is unclear, as a similar mouse LINE-1 probe gave no hybridization to duck DNA⁶, which should be closely related to chicken DNA.

It is not evident how this LINE-1 fragment could have a phenotypic effect on mouse cells. The mouse genome already contains thousands of copies of such sequences. Although LINES are derived from parental copies with two conserved open reading frames^{3,5}, the “*Blym-1*” specimen has many frameshifts in the first of these and could only translate a fragment of the second. (With regard to the biology of mouse LINES, it is interesting that the divergence of the “*Blym-1*” specimen from the specimen of ref. 5 is greatest at and around the junction between the two LINE-1 open reading frames.)

A human “*Blym-1* gene” was cloned by similar methods⁷. Homology was claimed between the chicken and human “*Blym-1*” genes, but it required many gaps and is not substantiated by diagonal matrix plots of the DNA sequences (not shown). “Human *Blym-1*” is unrelated to LINE-1 sequences, but the central part of it is a human *Alu* dispersed repeat, only 50 nucleotides of which were originally noticed⁷. This repeat covers nucleotides 262–567, with 74% homology to the consensus of ref. 8, and it substantially overlaps the proposed *Blym-1* open reading frame. Since *Alu* repeat sequences are species-specific, the significance of this gene also seems questionable.

JOHN ROGERS

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK

1. Cooper, G.M. & Neiman, P.E. *Nature* **287**, 656-659 (1980).
2. Goubin, G. *et al.* *Nature* **302**, 114-119 (1983).
3. Singer, M.F. & Skowronski, J. *Trends Biochem. Sci.* **10**, 119-122 (1985).
4. Hattori, M., Hidaka, S. & Sakaki, Y. *Nucleic Acids Res.* **13**, 7813-7827 (1985).
5. Loeb, D.D. *et al.* *Molec. cell. Biol.* **6**, 168-182 (1986).
6. Burton, F.H. *et al.* *J. molec. Biol.* **187**, 291 (1986).
7. Diamond, A., Devine, J.M. & Cooper, G.M. *Science* **225**, 516-519 (1984).
8. Deininger, P.L. *et al.* *J. molec. Biol.* **151**, 17-33 (1981).

SIR—The molecular clone designated chicken *Blym-1* was isolated from NIH 3T3 mouse cells transformed by DNA of a chicken B-cell lymphoma using sibselection to identify a clone which induced transformation of NIH 3T3 cells upon transfection¹. The cloned transforming gene hybridized to highly repeated sequences in mouse DNA and to a small family of related sequences in DNAs of chicken and human cells¹. Its nucleotide sequence was not closely related to any published sequences, although the predicted coding sequence was partially

homologous to transferrins¹. However, the transforming gene contained two short regions (20 nucleotides) which were 70–80% homologous to mouse satellite DNA and which were thought to account for the hybridization to repetitive mouse sequences¹. Therefore, the hybridization data did not determine whether the gene was of chicken or mouse origin¹. However, we considered it likely that the gene represented the transforming gene of chicken bursal lymphomas since serial transmission of transforming genes was established as a high-frequency event whereas activation of the transforming potential of normal cell genes was infrequent² and preliminary data indicated expression of homologous RNAs and a candidate protein product in chicken lymphoma as well as mouse cells.

Hattori *et al.*³ recently determined the complete sequence of a full-length member (TβG41) of the human long interspersed repetitive sequence family (LINES) and found that the TβG41 sequence was 60% homologous to that of chicken *Blym-1*. In addition, the amino-acid sequence predicted for the open reading frame of TβG41 was significantly homologous to transferrin³.

Loeb *et al.*⁴ have now reported the complete sequence of the A2 member of the mouse LINES family. The A2 sequence contains a region (nucleotides 2,402–2,990) which is 90% homologous to chicken *Blym-1*. These data indicate that chicken *Blym-1* is closely related to LINES and suggest that chicken *Blym-1* may represent a portion of a mouse LINES element whose transforming potential was activated during the process of molecular cloning from transformed NIH cells. Its activation might then be a consequence of truncation or rearrangement, as has been demonstrated for other transforming sequences^{5,7}.

To resolve the origin of chicken *Blym-1*, an oligonucleotide corresponding to positions 242–287 of the chicken *Blym-1* sequence¹ was synthesized and used as a probe for blot-hybridization analysis. This probe hybridized to mouse repetitive sequences but did not detect a homologous chicken sequence, although it hybridized to reconstruction digests containing the equivalent of one copy per genome of the chicken *Blym-1* clone. We therefore conclude that the transforming gene designated chicken *Blym-1* is a portion of a mouse LINES element whose transforming potential was probably activated during molecular cloning rather than the transforming gene originally detected by transfection of chicken bursal lymphoma DNAs⁸. We propose that the gene be redesignated mouse LINES transforming element (LINES-TE).

The mouse LINES-TE was used as a probe to isolate related sequences from a

genomic library of human Burkitt's lymphoma DNA⁹. One of these clones induced transformation of NIH 3T3 cells and was designated human *Blym-1*⁹. Human *Blym-1* has also been found to transform primary human foreskin fibroblasts to anchorage-independence¹⁰ and to induce growth of BALB/c 3T3 cells in PDGF-free medium (Fig. 1) using a modification of the co-selection assay described by Armelin *et al.*¹¹. Human *Blym-1* was also transferred to NIH cells transformed by genomic DNAs of Burkitt lymphoma cell lines⁹. Under stringent conditions, human *Blym-1* hybridizes to a unique sequence in human DNA⁹ and has been mapped to a unique chromosomal locus by *in situ* hybridization¹². The nucleotide sequence of human *Blym-1* indicated that it was overall 50–60% homologous to mouse LINES-TE and thus represented a related but divergent human gene which also shared partial homology to transferrin¹³. Consistent with the hybridization data, the nucleotide sequence of human *Blym-1* does not show significant homology to human LINES.

The LINES elements have been conserved throughout mammalian evolution and contain open reading frames which appear to have evolved under selection for protein function¹⁴. It is thus of interest that a portion of mouse LINES element (LINES-TE) can induce transformation and that a human transforming gene (*Blym-1*) is distantly related to the LINES family.

We thank Jacek Skowronski and Maxine Singer for pointing out the LINES homologies to us, and Charles D. Stiles for his development and help with the BALB 3T3 transformation assay.

GEOFFREY M. COOPER*
GERARD GOUBIN†
ALAN DIAMOND*
PAUL NEIMAN‡

*Dana-Farber Cancer Institute and
Department of Pathology,
Harvard Medical School,
Boston, Massachusetts 02115, USA

†Laboratoire d'Oncogenese Moleculaire,
Institut Curie, Paris, France

‡Fred Hutchinson Cancer Center,
Seattle, Washington 98104, USA

- Goubin, G., Goldman, D.S., Luce, J., Neiman, P.E. & Cooper, G.M. *Nature* **302**, 114–119 (1983).
- Cooper, G.M., Okenquist, S. & Silverman, L. *Nature* **284**, 418–421 (1980).
- Hattori, M., Hidaka, S. & Sakaki, Y. *Nucleic Acids Res.* **13**, 7813–7827 (1985).
- Loeb, D.D. *et al. Molec. cell. Biol.* **6**, 168–182 (1986).
- Muller, R. & Muller, D. *EMBO J.* **3**, 1121–1127 (1984).
- Molders, H., Defesche, D., Muller, T.I., Bonner, U.R. & Muller, R. *EMBO J.* **4**, 693–698 (1985).
- Takahashi, M., Ritz, J. & Cooper G.M. *Cell* **42**, 581–588 (1985).
- Cooper, G.M. & Neiman, P.E. *Nature* **287**, 656–659 (1980).
- Diamond, A., Cooper, G.M., Ritz, J. & Lane, M.-A. *Nature* **305**, 112–116 (1983).
- Sutherland, B.M., Bennett, P.V., Freeman, A.G., Moore, S.P. & Strickland, P.T. *Proc. Natn. Acad. Sci. U.S.A.* **82**, 2399–2403 (1985).

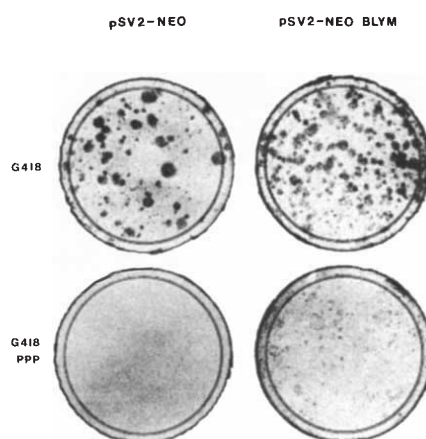


Fig. 1 Transformation of BALB 3T3 cells by human *Blym-1*. BALB/c 3T3 cells were transfected with 10 μ of pSV2neo (ref. 15) or pSV2neoBlym, which contains the 1-kilobase *EcoRI* fragment of pttuBlym-1 (ref. 13) inserted into the *EcoRI* site of pSV2neo. Recipient cells were subcultured 3 days after transfection into medium containing G418 (500 μ g ml⁻¹) and either 10% calf serum (designated G418) or 10% platelet-poor plasma (designated G418 PPP)¹¹. Plates were stained and photographed 14 days after transfection.

- Armelin, H.A. *et al. Nature* **310**, 655–660 (1984).
- Morton, C.C. *et al. Science* **223**, 173–175 (1984).
- Diamond, A., Devine, J. & Cooper, G.M. *Science* **225**, 516–519 (1984).
- Singer, M.F. & Skowronski, J. *Trends biochem. Sci.* **10**, 119–122 (1985).
- Southern, P.J. & Berg, P. *J. molec. appl. Genet.* **1**, 327–341 (1982).

Imanishi and Halstead: intraspecific competition?

SIR—Halstead has done a service to readers of *Nature* by bringing to their attention¹ the work of Imanishi, the Japanese evolutionist. His treatment of the scientific basis of Imanishi's theory is, however, unbalanced and, perhaps unconsciously, a tad culturally biased. He states, for example:

Imanishi's evolutionary theory is a poetic vision, it is beautiful to contemplate but it is a dream and it is Japanese in its unreality . . . unhappily it has no place in the scientific understanding of the real world.

Halstead implies that there is a particular unreality peculiar to the Japanese culture, but it is the second half of the quotation on which I wish to comment. What is the essence of Imanishi's "poetic vision", and does it deserve such harsh criticism? Halstead informs us that the theory minimizes the role of competition in community dynamics (based in Imanishi's field work on "habitat segregation") as well as in speciation. The group receives greater emphasis than the individual; harmony is emphasized relative to struggle.

In rebutting the scientific basis of Imanishi's theory Halstead deals briefly with two aspects of the theory: (1) the role of co-operation in evolution; and (2) the accumulated evidence on the role of interspecific competition in community dy-

namics. Both treatments are superficial at best and do not reflect the current turmoil in the scientific literature on both the role of interspecific competition in habitat partitioning and that of natural selection in speciation. For example, Halstead, in relation to evidence for the importance of competition at the community level, cites Schoener² but fails to indicate that a subsequent issue of *American Naturalist* was devoted to a contentious debate on this very question. Imanishi no doubt would not deny that interspecific competition exists (as Schoener documents), but the real issue is whether this process is critical to community dynamics or speciation. Even a cursory glance at the recent ecological literature is sufficient to conclude that there is far from a consensus.

More critically, perhaps, Halstead appears to have confused the roles of intraspecific and interspecific competition in the evolutionary synthesis. His confusion leads to unjustified characterization of Imanishi's theory. Halstead states, without substantiation, that many other evolutionary theories reject the role of intraspecific competition. He then states that the novelty of Imanishi's theory is the rejection of interspecific competition as well. As August Weismann³, the architect of neo-darwinism, so clearly states in the centenary volume of Darwin's birth:

The 'struggle for existence', which Darwin regarded as taking the place of the human breeder in free nature, is not a direct struggle between carnivores and their prey, but is the assumed competition for survival between individuals of the same species . . . (emphasis in original)

Natural selection within darwinism and neo-darwinism required intraspecific competition. Surely not just Imanishi's but any theory (I wish Halstead had been more specific on these other theories) that rejects the role of such competition in evolution is non-darwinian.

In sum, Halstead's evaluation of the role of competition in speciation in relation to both Imanishi's theory of evolution and the evolutionary synthesis is misleading. It is unfortunate that his more-detailed critique of Imanishi's work is in Japanese and thus will have restricted critical review. Given the contents of the *Nature* article, however, the Japanese research community should not consider Halstead's position as necessarily being representative of the present literature on ecology and evolutionary biology.

M. SINCLAIR
Invertebrates and Marine Plants Division,
Fisheries Research Branch Scotia-Fundy
Region,
PO Box 550, Halifax,
Nova Scotia B3J 2S7, Canada

- Halstead, B. *Nature* **317**, 587–589 (1985).
- Schoener, T.W. *Am. Nat.* **122**, 240–285 (1983).
- Weismann, A. in *Darwin and Modern Science* (ed. Seward, A.C.) 18–65 (Cambridge University Press, 1909).

Taxonomy and control of *Salvinia molesta*

from P. A. Thomas and P. M. Room

Outside of its native South America, salvinia is one of the world's worst aquatic weeds. Conventional control is prohibitively expensive, but now biological control using a beetle has become a reality.

THE plant *Salvinia molesta* Mitchell, generally known as salvinia, is a free-floating aquatic fern indigenous to south-east Brazil¹. It has some attraction as a botanical curiosity and aquarium plant; during the past 50 years, it has been spread widely through the tropics. It almost certainly arrived in Asia, for example, via European botanical gardens^{2,3}. The plant is sterile, but undergoes rapid vegetative propagation by growth and fragmentation⁴. Doubling in size in as little as 2.2 days⁵, it can quickly cover lakes and slow-moving rivers with mats up to 1 m thick. As a result, salvinia is generally ranked with the equally intransigent water hyacinth as one of the two worst aquatic weeds in the world.

Thick mats of salvinia prevent the passage of large diesel-powered boats and even a single layer of plants is a major obstacle to paddle-powered canoes. As a result, transport by water and commercial and recreational fishing can be severely impeded or halted. Mats block access to drinking water by humans, domestic stock and wildlife, clog irrigation and drainage canals and, during floods, sweep fences and other light structures before them. The plant is a major weed of rice and harbours the hosts of human diseases such as schistosomiasis^{6,7}. By cutting off light to submerged plants, salvinia mats depress oxygen concentrations and increase those of carbon dioxide and hydrogen sulphide in the water beneath⁸, which can lead to extinction of most of the benthic fauna⁹.

Nowhere have the adverse socio-economic effects of salvinia been more serious than in the floodplain of the Sepik River in Papua New Guinea. A few plants were probably introduced via an aquarium in 1971 or 1972^{10,11}; by the early 1980s, they had multiplied to infest all the 500 km² of lakes in the lower half of the floodplain, completely covering 250 km² of water surface. Salvinia became a major threat to the existence of the 80,000 people living in the area, which is almost devoid of roads due to annual flooding and extensive swamps and waterways. The staple carbohydrate is obtained from the pith of sago palms and the use of canoes is vital for harvesting the trunks and towing them from swamps to villages. The movement of canoes is also essential for fishing, for

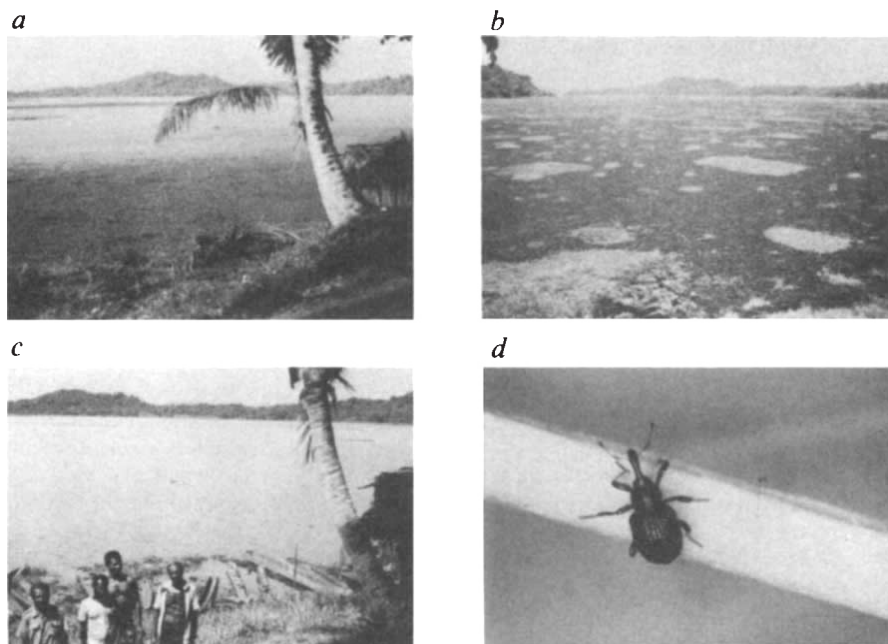


Fig. 1 Lake Kabufwe in Papua New Guinea. *a*, In October 1983 when access to this village by canoe had been halted by salvinia before *C. salviniae* was having any effect; *b*, in June 1984 when the beetles were killing the salvinia (the patches are grasses and other vegetation which had colonized the floating salvinia); *c*, in November 1984 when almost all salvinia had been destroyed and the village could again be reached by canoe; *d*, an adult *Cyrtobagous salviniae* on a matchstick.

access to markets, schools and medical assistance.

Early control

The first attempts to control salvinia were made in Sri Lanka, where the weed became established in the 1940s¹². High volume application of emulsifiable oils containing pentachlorophenol was used to kill the plant which had infested 10,000 hectares of rice paddy and 900 hectares of waterways¹³. Salvinia gained even more notoriety when, in 1959, it invaded Lake Kariba on the Zambia/Zimbabwe border. At its peak in 1962, the weed covered 1,000 km², or 21%, of what was then the world's largest man-made reservoir⁸. Of 61 chemicals screened for herbicidal activity against salvinia, the two most effective were paraquat and sodium arsenite (E. R. Hattingh, unpublished reports to the Lake Kariba Coordinating Committee, 1962, 1963). Use of the latter was ruled out, at the rates required for weed control, by its toxicity to all forms

of life. Field trials demonstrated the efficacy of paraquat, but no control programme was undertaken because costs were prohibitive.

Paraquat has, nevertheless, been used to provide short-term control of salvinia in Kenya¹⁴, Sri Lanka¹⁵, Botswana¹⁶, Australia¹⁷ and Papua New Guinea¹⁸. A number of other chemicals such as Karmex W¹⁹, anhydrous ammonia²⁰, formalin²¹, the molluscicide niclosamide²² and the fungicide thiram²³ have also been shown to have herbicidal activity against the plant, but have been used only in trials. AF101, a mixture of kerosene and detergent developed in Australia²⁴, gives rapid kill of salvinia and has been used with short-term success in field operations¹⁷, but it is now clear that claims of the chemical eradication of salvinia^{13,14} were premature.

Despite the availability of herbicides which kill the weed, attempts to control salvinia chemically have, without exception, failed to provide a long-term solu-

tion. The reasons for this are basically economic. Regrowth from the inevitable survivors of a spraying programme is exponential until the density of the weed again approaches the level that provoked spraying. At Lake Moondarra, Australia, use of AF101 sprayed from a hovercraft and a helicopter was abandoned in December 1978 after \$A 160,000 had been spent⁵. At the time, the plant was doubling in size in less than 3 days, implying the need to kill more than half the total infestation every 3 days if control were to be achieved. In the Sepik floodplain, it was estimated that to reduce and keep the salvinia infestation to less than 10% of the water surface would require an initial outlay of \$1 million followed indefinitely by an annual \$500,000 (ref. 18). [Unless otherwise indicated, as for example by \$A, dollar amounts refer to US dollars.]

Economic constraints are also the main reason for the failure of physical or mechanical control measures. Manual removal from small water bodies can be useful in the early stages of an infestation, but once the weed is established, biomasses of about 80 tonnes/ha fresh weight and the potential for rapid regrowth make this impractical. The use of weed-harvesting machines was considered in Australia, but even in winter, when salvinia doubling times are as long as 40–60 days, the capacity of large infestations for regrowth exceeded the removal capacity of the machines²⁵. As with water hyacinth, mechanical harvesting is not cost-competitive with chemical control²⁶.

Floating booms have some value in confining salvinia and maintaining adjacent waters in a relatively weed-free condition, but are subject to enormous stresses by windblown salvinia mats. On Lake Kariba, a boom slung on a 5 cm diameter steel cable across the Sebungwe Narrows, where weed was entering the western end of the lake, was abandoned after it had been broken three times (P.A.T., unpublished data). In Papua New Guinea, fishing waters were successfully kept clear for short periods by booms up to 750 m long built from logs tied to 13-mm steel cables, but waterlogged timbers required regular replacement and anchor points on the swampy banks frequently gave way¹⁸.

More positively, some attention has been given to making use of the huge biomasses provided by salvinia. The plant has been employed as a compost and mulch^{27,28} and in Asia is used, together with other aquatic weeds, to supplement livestock fodder^{29,30}. There have been investigations of its use for the treatment of sewage effluent³¹, paper making³² and the generation of biogas³³, but none has led to large-scale utilization, no doubt because of the cost of moving the water, which is 96% of the fresh weight. For such schemes to be viable, a sustained supply of raw material must be assured, which is

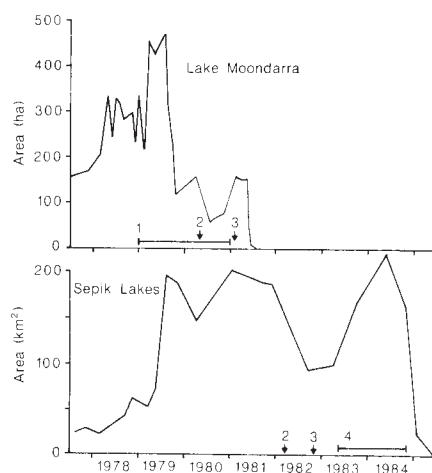


Fig. 2 Changes in the area covered by salvinia on L. Moondarra (above) and on 100 lakes in the Sepik floodplain (below). (1) period of drought which reduced lake area by 55%; (2) *C. salviniae* first introduced (into field cages); (3) beetles released from cages; (4) period of manual distribution of beetles throughout salvinia infestations in the Sepik floodplain.

incompatible with the concept of control.

The underlying difficulty of salvinia control is that, as well as the potentially undesirable environmental consequences of the long-term use of herbicides, all measures entail repeated and often expensive inputs yet cannot maintain salvinia populations below nuisance levels. In retrospect, only biological control could have been effective.

Biological control

The possibility of the biological control of salvinia was prompted by the appearance of the weed on Lake Kariba in 1959. At that time, the plant had been misidentified as *Salvinia auriculata* Aubl³⁴. Accordingly, when the Commonwealth Institute of Biological Control (CIBC) was commissioned to search for potential biological control agents in South America, collections were made from *Salvinia auriculata* (*sensu strictu*), which was widespread there and which appeared to be the same as that which had by then become known as 'Kariba weed', and which was starting to make its presence felt in several tropical countries. Three insects collected in northern S. America were recommended as potential control agents^{35,36}, a moth, *Samea multiplicalis* Guenée, a grasshopper, *Paulinia acuminata* (De Geer) and a beetle, *Cyrtobagous singularis* Hustache.

During the 1960s and 1970s, releases of all three insects were made in various countries. *C. singularis* was released on Lake Kariba³⁷, in Fiji³⁸ and Botswana¹⁶, but the beetle was established only in Botswana³⁹, where it appears to have had no effect on salvinia populations. *S. multiplicalis* failed on Lake Kariba⁴⁰ and, although established in Fiji³⁸ has not pro-

vided control. *P. acuminata* appeared to establish in Botswana¹⁶, but subsequently seems to have died out (D.L.C. Procter, personal communication). The grasshopper also failed in Kenya and Sri Lanka⁴⁰, but has been established in Fiji³⁸, India⁴⁰ and on Lake Kariba³⁷, where there is some circumstantial^{37,42} and controversial evidence that the insects may have provided some degree of control.

Where salvinia is a weed, a few native insects have been recorded feeding on the plant. Some, such as the cosmopolitan aphid, *Rhopalosiphum nymphae* (Linn), and the Pyralid moth, *Nymphula responsalis* Wlk, both native to South-East Asia and Australia, have been investigated as possible biological control agents^{43,44}. Augmenting the effects of native natural enemies, or incorporating them into integrated control programmes, is attractive, but these studies have not progressed beyond the experimental stage. The aquatic snail *Marisa cornuarietis* L. eats *Salvinia rotundifolia* Willd⁴⁵, but also attacks rice seedlings and so is unsuitable as a control agent⁴⁶. Fungal infections which occasionally appear in association with moribund salvinia are almost certainly secondary saprophytes. Laboratory isolates of an apparently parasitic *Spicariopsis* sp., collected from salvinia on Lake Kariba, failed to show primary pathogenic features in laboratory experiments⁴⁷.

CSIRO's breakthrough

By 1970 it had been realized that the 'Kariba Weed' was not *S. auriculata* but a species previously undescribed. At the same time, there was found in South America a herbarium specimen of a plant identical with 'Kariba Weed' which had been collected, together with two other species of *Salvinia*, from the Rio de Janeiro botanic garden in 1941. Evidence in support of the suggestion that the new species might be a horticultural hybrid⁴⁸ included its sterility, its size and aggressive growth (? hybrid vigour) and the fact that it appeared to be pentaploid⁴⁹. It was described as *Salvinia molesta* D. S. Mitchell in 1972; eventually, in 1978, the native range of *S. molesta* was discovered in S. E. Brazil^{50,51}.

Collections of insects in the native range of *S. molesta* revealed that, as with previous collections from *S. auriculata*³⁶, the three most promising candidates for biological control were *Paulinia acuminata*, *Samea multiplicalis* and what appeared to be *Cyrtobagous singularis*⁵². After further studies and host-specificity testing⁵³, the last of these was released on Lake Moondarra in Australia and when, in 14 months, it had destroyed a 200-hectare salvinia infestation (Fig. 2), it was assumed that a biotype of *C. singularis* adapted to *S. molesta* had been discovered⁵⁴. But subsequent investigations showed the successful insect to be a previously unde-

scribed species, now known as *Cyrtobagous salviniae* Calder and Sands⁵⁵ (Fig. 2d). The same insect has controlled salvinia elsewhere in Australia⁵⁶, in Papua New Guinea^{18,57} and is rapidly achieving control in Namibia (ref. 58 and I. W. Forno, personal communication) and India (P. J. Joy, personal communication).

In Papua New Guinea, *C. salviniae* has provided the most spectacular example so far of biological control of an aquatic weed (Figs. 1, 2). In September 1982, the beetle was established in a Sepik lake, but only when the nitrogen content of salvinia in field cages was raised by regular applications of urea⁵⁹. Fortunately, the insects continued to multiply when large numbers were released from the cages onto an unfertilised mat of weed. There is some evidence that damage caused by the beetles increases nitrogen levels in the tissues remaining, with the result that, at low ambient nitrogen levels, a critical density of insects may be necessary to tip the nutritional balance in favour of establishment⁵⁹. Subsequent field and laboratory studies have confirmed the importance of nitrogen levels, both for *C. salviniae* and for *Samea multiplicalis*⁶⁰⁻⁶².

Large-scale redistribution of beetles in the Sepik infestation began in May 1983 (Fig. 2), 3-4 generations after establishment outside the cages. Infested salvinia containing all stages of the beetle was carried in jute sacks to new release sites. By the end of 1984, about 850,000 adults and an unknown number of eggs, larvae and pupae, had been distributed among 130 lakes. Natural dispersal and deliberate movement of infested weed by villagers had spread beetles throughout the rest of the floodplain infested with salvinia. The pattern of attack was essentially the same in all lakes: it numbers of beetles and the damage they caused increased dramatically about 6 months after their introduction (Fig. 3).

By August 1985, the salvinia cover had been reduced from ca. 250 km² to 2 km² and *C. salviniae* had destroyed 2 million tonnes of the weed in two years. Although some salvinia was flushed down river as mats broke up, most died and sank *in situ*. There have been no indications of any harmful side effects from this process. In all cleared lakes, a residual population of beetles on scattered plants growing among marginal vegetation appears to be holding the weed in check. In the earliest cleared lagoons, this balance has now persisted for more than 2 years.

The success of *C. salviniae* contrasts with the failures of the other salvinia insects, particularly the very similar *C. singularis*. Adult *C. salviniae* feed selectively on growing points, which must be of considerable significance to a plant which relies on vegetative propagation⁶³. The larvae tunnel through the rhizome, destroying vascular tissues. Adult *C. sin-*

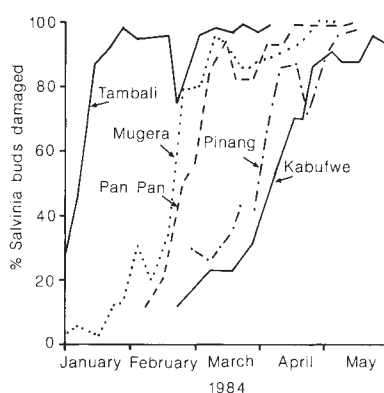


Fig. 3 Rapid increases in damage caused by *C. salviniae* to salvinia on five lakes in the Sepik floodplain in early 1984. In each case, damage to buds reached 50% about 7 months after beetles were introduced and most of the salvinia was destroyed in the following 5 months.

gularis, by contrast, feed more at random, allowing more buds to survive, and their larvae feed externally without destroying vascular tissues^{60,61}. Even so, the crucial difference is probably related to the fact that populations of *C. salviniae* increase until the food supply runs out, while populations of *C. singularis* seem to stop increasing much earlier, as in southern Africa in the 10 years since *C. singularis* was released³⁹.

There appear to be no close ecological homologues of *Cyrtobagous* spp. outside S. America and it is possible that there are no natural enemies preadapted to attack them in a density-dependent fashion outside their native ranges. No parasites or diseases of *Cyrtobagous* spp. are known. This contrasts with *S. multiplicalis*, which has become a host for at least three hymenopterous parasites and one protozoan disease in the four years it has been in Australia (I. W. Forno, personal communication).

The correct identification of *S. molesta* and the discovery of its native range were clearly the keys to biological control of salvinia. This lends support to the idea that successful control agents are most likely to be collected from the same species as the target pest⁶⁴. Nevertheless, the control of *Opuntia inermis* and *O. stricta* prickly pears in Australia by *Cactoblastis cactorum* (Berg), which never encountered the plants in its native Argentina, remains a powerful counter example⁶⁵.

Future prospects

Since *S. molesta* is asexual, the world's entire population of the plant is ostensibly genetically identical. The lack of variation suggests that, other things being equal, the susceptibility of the weed to *C. salviniae* should not differ from place to place.

Of the factors which may effect the success of the insect, the most important is likely to be temperature. *C. salviniae* eggs

will not hatch below 19 °C (ref. 53). Since salvinia will continue to grow at 10 °C (ref. 66) and can survive light frosts, it is clear that the plant will survive in areas which, for part of the year at least, will not support *C. salviniae*. But there are few serious salvinia infestations outside the tropics, so that climatic constraints should not be an impediment to wider use of the beetle against most infestations.

Low nitrogen levels may impede attempts to establish beetles, particularly as first introductions often involve small populations. Application of an appropriate fertilizer at the release site could increase nutrient levels until the critical population density had been reached. Evidence from Papua New Guinea also suggests that low nutrient levels may be associated with thick weed mats⁶⁷; the failure of *C. salviniae* to establish under these conditions on one Sepik lake was overcome by destroying part of the weed mat with paraquat, which thinned out the surviving plants and appeared to catalyse an immediate population explosion of beetles.

The existence of pre-adapted predators, parasites and diseases of a control agent outside its native range cannot be predicted⁶⁸, but the unusual growth form of salvinia, the specialized habits of *C. salviniae* and the apparent lack of natural enemies encountered so far suggest that it is unlikely that natural enemies of the beetle will interfere with future attempts to use it as a control agent.

Building on the work already carried out, the cost of introducing *C. salviniae* into each additional country is likely to range between \$100,000 and \$500,000 over 2 to 3 years, depending on whether several countries in one region can be serviced together and on whether they require that further host-specificity tests should be carried out. The results of host-specificity tests already performed, and field experience in the countries where *C. salviniae* has been released, all indicate that the beetle will feed only on plants in the genus *Salvinia*, but many countries are understandably sensitive to the slightest potential threat to their key crops and may require tests to confirm that those crops will not be at risk.

Conclusions

C. salviniae now provides cost-effective, environmentally sound and apparently permanent biological control of salvinia in Australia and Papua New Guinea. Progress is good in India and Namibia. Alternative methods of control, such as the use of herbicides or physical removal, are not economically viable in comparison because they must be applied repeatedly and indefinitely. *C. salviniae* is likely to prove equally effective in other tropical countries having salvinia infestations, such as Sri Lanka, Indonesia and Malay-

sia, and within a few years one of the world's worst weeds should cease to be a problem.

The success we have reconstructed vividly demonstrates the crucial importance of sound taxonomy, both of plants and animals, for understanding and solving ecological problems. □

P. A. Thomas was project coordinator of the PNG/UNDP/FAO Salvinia Control Project, PO Box 495, Wewak, Papua New Guinea; present address: 8 Sadlers Court, Mattock Way, Abingdon OX14 2PA, UK. P. M. Room is at the CSIRO Division of Entomology, Private Bag No. 3, Indooroopilly, Queensland 4068, Australia.

1. Forno, I. W. *Aquat. Bot.* **17**, 71-83 (1983).
2. Cook, C. D. K. & Gut, B. J. *Pest Artic. News Summ.* **17**, 438-447 (1971).
3. Cook, C. D. K. *J. aquat. Pl. Mgmt* **23**, 1-6 (1985).
4. Room, P. M. *J. Ecol.* **71**, 349-365 (1983).
5. Farrell, T. P. *Proc. Aust. Wat. Resources Coun. Semin.*, Canberra, 57-71 (Department of National Development, Canberra, 1979).
6. Chow, C. Y., Thevasagayam, E. S. & Wambeck, E. G. *Bull. Wild Hlth Org.* **12**, 365-369 (1955).
7. Hira, P. R. *Nature* **224**, 670-672 (1969).
8. Mitchell, D. S. *Hydrobiologia* **34**, 448-464 (1969).
9. McLachlan, A. J. *Archs Hydrobiol.* **66**, 212-231 (1969).
10. Gewertz, D. B. *Sepik River Societies* (Yale University Press, New Haven, 1983).
11. Mitchell, D. S., Petr, T. & Viner, A. B. *Envir. Conserv.* **7**, 115-122 (1980).
12. Senaratna, J. E. *Trop. Agric. Mag. Ceylon agric. Soc.* **99**, 146-149 (1943).
13. Williams, R. H. *Trop. Agric. Trin.* **33**, 145-157 (1956).
14. Plant Protection Information Service, ICI Plant Protection Limited (1965).
15. Dias, G. R. W. *Wild Crops* **19**, 64-68 (1967).
16. Edwards, D. & Thomas, P. A. *Proc. 2nd natn. Weeds Conf.*

- S. Afr. Stellenbosch Univ.* **221-237** (Balkema, Cape Town, 1977).
17. Farrell, T. P. *Proc. 1st Conf. Coun. Aust. Weed Sci. Soc.* 179-187 (Council of Australian Weed Science Societies, Melbourne, 1978).
18. Thomas, P. A. *Pl. Prot. Bull. F.A.O.* **32**, 50-56 (1985).
19. George, K. *Trans. Bose Res. Inst.* **21**, 1-8 (1956).
20. Aryaratne, J. K. P. *Proc. 6th Asian-Pacific Weed Sci. Soc. Conf.* 282-284 (1977).
21. Begg, G. W. *Newsl. Linnol. Soc. S. Afr.* **17**, 22-26 (1971).
22. Wild, H. & Mitchell, D. S. *Pflschut-Nachr. Bayer* **23**, 105-110 (1970).
23. Webster, R. H. *Pest Artic. News Summ.* **23**, 333-334 (1977).
24. Diatloff, G., Lee, A. N. & Anderson, T. M. *J. aquat. Pl. Mgmt* **17**, 24-27 (1979).
25. Mitchell, D. S. *The Incidence and Management of Salvinia molesta in Papua New Guinea* (Office of Environment & Conservation, Waigani, Papua New Guinea, 1979).
26. Canellos, G. *Aquatic Plants and Mechanical Methods for Their Control* (Mitre Corporation, McLean, Virginia, 1981).
27. Boyd, C. E. in *Aquatic Vegetation and its Use and Control* (ed. Mitchell, D. S.) 107-115 (Unesco, Paris, 1974).
28. Subagyo, T. & Vuong, N. V. *BIOTROP Newsl.* **12**, (1975).
29. Anon. in *Making Aquatic Weeds Useful. Some perspectives for Developing Countries* (eds Ruskin, F. R. & Shipley, D. W.) 56-64 (National Academy of Sciences, Washington, D.C., 1976).
30. Little, E. C. S. *Handbook of Utilization of Aquatic Plants* (FAO, Rome, 1968).
31. Finlayson, C. M., Farrell, T. P. & Griffiths, D. J. *Water Res. Fdn Aust. Rep. No. 57* (1982).
32. Bhambe, S. & Bhardwaj, K. R. *Hydrobiologia* **65**, 209-211 (1979).
33. *Aquaphyte 3(2)* (ed. Ramey, V.) 67-73 (University of Florida, Gainesville, Florida, 1983).
34. Schelpe, E. A. C. L. E. J. S. *Afr. Bot.* **27**, 181-187 (1961).
35. Bennett, F. D. *Feeding Tests with Insects Attacking Salvinia auriculata undertaken at Belem during October-December 1961* (CIBC Report, 1964).
36. Bennett, F. D. *Proc. Sthn. Weed Conf.* **19**, 497-504 (1966).
37. Mitchell, D. S. & Rose, D. J. W. *Pest Artic. News Summ.* **25**, 171-177 (1979).
38. Kamath, M. K. *Fiji agric. J.* **41**, 55-72 (1979).
39. Procter, D. L. C. *Botswana Not. Rec.* **15**, 99-101 (1983).
40. Bennett, F. D. *Proc. Symp. Water Quality Management through Biol. Control* 28-35 (University of Florida, Gainesville, 1975).

41. Chisolm, I. F. *Bull. ent. Res.* **69**, 111-114 (1979).
42. Marshall, B. E. & Junor, F. J. R. *Hydrobiologia* **83**, 477-484 (1981).
43. John, K. C. & Nair, N. B. *Entomol.* **7**, 381-384 (1982).
44. Kam-Wing, L. & Furtado, J. I. *Hydrobiologia* **56**, 49-61 (1977).
45. Seaman, D. E. & Porterfield, W. A. *Weeds* **12**, 87-92 (1964).
46. Bennett, F. D. in *Aquatic Vegetation and its Use and Control* (ed. Mitchell, D. S.) 99-106 (Unesco, Paris, 1974).
47. Loveless, A. R. *Ann. appl. Biol.* **63**, 61-69 (1969).
48. Mitchell, D. S. & Thomas, P. A. *Tech. Pap. Hydrol. No. 12* (Unesco, Paris, 1972).
49. Loyal, D. S. & Grewal, R. K. *Cytologia* **31**, 330-338 (1966).
50. Mitchell, D. S. *Br. Fern. Gaz.* **10**, 251-252 (1972).
51. Forno, I. W. & Harley, K. L. S. *Aquat. Bot.* **6**, 185-187 (1979).
52. Forno, I. W. *Proc. V. Int. Symp. Biol. Control Weeds*, Brisbane 1980, 167-173 (CSIRO, Melbourne, 1981).
53. Forno, I. W., Sands, D. P. A. & Sexton, W. *Bull. ent. Res.* **73**, 85-95 (1983).
54. Room, P. M., Harley, K. L. S., Forno, I. W. & Sands, D. P. A. *Nature* **294**, 78-80 (1981).
55. Calder, A. A. & Sands, D. P. A. *J. Aust. ent. Soc.* **24**, 57-64 (1985).
56. Room, P. M., Forno, I. W. & Taylor, M. F. J. *Bull. ent. Res.* **74**, 505-516 (1984).
57. Thomas, P. A. & Room, P. M. *Proc. VI Int. Symp. Biol. Control Weeds*, Vancouver, 567-574 (1985).
58. Schlettwein, C. H. G. & Hamman, P. F. S. W. *Afr. Ann. Rep.* **1984**, 49-51 (1985).
59. Room, P. M. & Thomas, P. A. *J. appl. Ecol.* **22**, 139-156 (1985).
60. Sands, D. P. A. & Schotz, M. *Proc. VI Int. Symp. Biol. Control Weeds*, Vancouver, 1984, 551-556 (1985).
61. Sands, D. P. A., Schotz, M. & Bourne, A. S. *Entomologia exp. appl.* **34**, 291-296 (1983).
62. Taylor, M. F. J. *J. Insect Physiol.* **30**, 779-785 (1984).
63. Room, P. M. *Proc. VI Int. Symp. Biol. Control Weeds*, Vancouver, 1984, 89-102 (1985).
64. Wapshere, A. J. *Pest Artic. News Summ.* **21**, 295-303 (1975).
65. Mann, J. *Bull. U.S. natn. Mus.* **256**, 1-158 (1969).
66. Cary, P. R. & Weerts, P. G. J. *Proc. V. Int. Symp. Biol. Control Weeds*, Brisbane 1980, 35-46 (CSIRO, Melbourne, 1981).
67. Room, P. M. & Thomas, P. A. *Aquat. Bot.* (in the press).
68. Goeden, R. D. & Louda, S. M. A. *Rev. Ent.* **21**, 325-342 (1976).

REVIEW ARTICLE

Cachectin and tumour necrosis factor as two sides of the same biological coin

Bruce Beutler & Anthony Cerami

Laboratory of Medical Biochemistry, The Rockefeller University, New York, New York 10021, USA

In response to invasive stimuli, macrophages secrete cachectin, a multipotent protein. Prominent among its biological effects is the ability to induce wasting (cachexia) as well as a lethal state of shock. The identity of cachectin and tumour necrosis factor has led to a new view of its therapeutic potential.

INVASIVE diseases, whether infectious or neoplastic, frequently precipitate metabolic derangements that threaten the integrity of the host. Abnormalities of host glucose, protein and lipid metabolism are well documented in acute and chronic disease states¹⁻³, and represent systemic consequences of defective cellular homeostasis. Gross imbalances of host physiology, manifested as a wasting of the body (cachexia) or shock, may of themselves lead to death⁴.

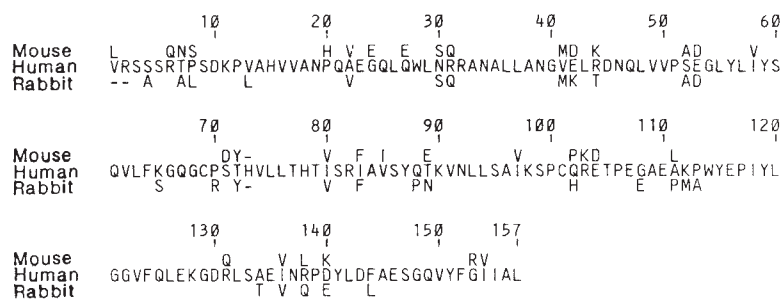
In attempting to explain these pathological states, much emphasis has been placed on the invasive stimulus (for example, the infectious agent or tumour) as the proximal cause of the problem. It has been widely supposed, for example, that microbial pathogens, or their products, are directly responsible for shock, and for the wasting associated with infection. Recent observations have challenged this view. Both acute and chronic metabolic derangements seem to be mediated largely by the

immune system. In response to a variety of invasive stimuli, reticuloendothelial cells and lymphocytes secrete cytokines that are capable of altering host metabolism. These cytokines include interleukin-1 (IL-1)⁵⁻⁷, lymphotoxin⁸⁻¹¹, γ -interferon¹², and the substance known as either cachectin or tumour necrosis factor¹³⁻¹⁵. Cachectin is of particular interest in that it has been implicated as important in inducing shock¹⁶ and cachexia¹⁷⁻²⁰ in animal models, but only recently has it been isolated and characterized by investigators studying both cachexia and the necrosis of tumours.

Identification and isolation

Coley^{21,22} was the first to describe the phenomenon of tumour necrosis, as it occurred in human cancer patients challenged with various poorly defined bacterial toxins, derived from *Streptococcus* and *Serratia*. Subsequently, through the work of Shear

Fig. 1 Primary structure of mouse⁵⁰, human^{15,43,44} and rabbit⁵² cachectin (TNF). Only the mature polypeptide hormones (devoid of propeptide sequence) are shown. Thirty-four substitutions distinguish the rabbit from the mouse protein, while there are 31 differences between the rabbit and human proteins, and 34 differences between human and mouse. Among the three species there is a total of 45 mutable residues. The single 'most conserved' region is that spanning residues 114–130; this segment is invariant among the three species. The region spanning residues 45–69 is also relatively conserved. Both areas exhibit a high degree of homology to corresponding regions of the lymphotoxin molecule¹¹. Cys 69 is thought to be disulphide-bonded to Cys 101 in cachectin (TNF) from all three species⁴².



and others^{23,24}, it was established that endotoxin (lipopolysaccharide, LPS) was particularly effective at inducing tumour necrosis. In 1975, Carswell *et al.* reported the existence of a factor in the serum of BCG-primed, endotoxin-treated animals which was able to elicit haemorrhagic necrosis of tumours in recipient animals²⁵; the active principle was referred to as 'tumour necrosis factor' (TNF). A variety of similar observations, using other experimental models, followed^{26–31}.

Several lines of evidence suggested that the macrophage was the principal source of this factor, and its production by isolated macrophages was eventually demonstrated^{26,27,29,30,32–39}. Serum obtained from primed, endotoxin-treated animals was found capable of selectively lysing some (but by no means all) tumorigenic cells, *in vitro* as well as *in vivo*^{40,41}. Although the mechanism underlying this selective cytolytic activity has yet to be clarified, interest in TNF has been considerable on account of its possible use in the treatment of neoplastic diseases. Human TNF was finally purified to homogeneity and partially sequenced by Aggarwal *et al.*⁴². The full sequence followed after the cDNA of TNF was cloned from a cDNA library, using a synthetic oligonucleotide probe based on the established N-terminal sequence of the protein¹⁵. Several other investigators, working independently, reported sequences that were identical, or nearly so^{43,44}.

While these studies of endotoxin-induced necrosis of tumours were in progress, we were attempting to delineate basic mechanisms of cachexia, the wasting associated with many chronic disease states. Rouzer *et al.*⁴⁵ observed that rabbits infected with *Trypanosoma brucei* become profoundly cachectic, losing up to 50% of their body mass in the course of the disease, despite extremely low levels of parasitaemia and negligible involvement of parenchymal organs. In the final stages of the infection, the animals, though severely wasted, become remarkably lipaemic; triglycerides are the principal components of the raised serum lipid and there is an associated systemic deficiency of the enzyme lipoprotein lipase. It was postulated that a host factor, produced in response to the infection, is responsible for this clearly defined biochemical abnormality.

It was noted that whereas lipoprotein lipase activity became suppressed in endotoxin-sensitive (C3H/HeN) mice following LPS administration, endotoxin-resistant (C3H/HeJ) animals did not respond to LPS in this way. However, a factor in the serum of LPS-injected C3H/HeN mice acted to suppress lipoprotein lipase activity in endotoxin-resistant animals¹⁷. This factor was termed 'cachectin' in view of its presumed involvement in the pathogenesis of cachexia. Subsequent experiments showed that cachectin was produced *in vitro* by macrophages in response to LPS¹⁸ and to other substances of microbial and parasitic origin¹⁹. More importantly, the observation that cachectin was able completely to suppress lipoprotein lipase activity in isolated adipocytes (3T3-L1 cells)^{18,46,47}, became the basis of a sensitive bioassay that was used to purify the hormone¹³.

On determining the N-terminal sequence of mouse cachectin, Beutler *et al.*¹⁴ found that it was strikingly similar to that of human TNF. It was also observed that highly purified cachectin

possessed potent TNF activity *in vitro*, suggesting that 'cachectin' and 'TNF' are homologous molecules. The complete primary structure of both mouse and human proteins, which has now been determined by DNA sequencing^{48–50}, confirms that cachectin and TNF bioactivities both derive from a single, highly conserved protein.

At present, we believe that the term cachectin (derived from cachexia, which in turn comes from the Greek *kakos* (bad) plus *exein* (state or habit of body)) is a better term than tumour necrosis factor, since it defines the broader biological properties of this important protein. We will therefore use the term cachectin throughout the remainder of this review.

Biosynthesis and structure

Cachectin is a polypeptide hormone composed of subunits of relative molecular mass (M_r) 17,000 (17K) arranged in dimeric, trimeric or pentameric form depending on species and method of isolation^{13,14,42,51}. It is unclear whether a multimeric arrangement is required for biological activity. The protein may be eluted from denaturing gels in monomeric form with some recovery of bioactivity¹³. In addition, it can be renatured following treatment with 6 M urea under reducing conditions. Although cachectin bioactivity can withstand temperatures exceeding 50 °C, the protein is irreversibly denatured by boiling. There is no firm evidence to suggest that the protein is glycosylated in any species, although the mouse protein sequence contains potential glycosylation sites^{48–50}.

The primary structure of cachectin (TNF) derived from three mammalian species (rabbit⁵², mouse^{48–50} and human^{15,43,44}) has been determined to date (Fig. 1). Compared with the mature mouse protein, which is 156 residues long, the rabbit protein lacks two N-terminal residues and is therefore 154 residues long, while the human protein consists of 157 amino acids, owing to the insertion of a histidine at position 73.

Both the mouse and human proteins (and presumably the rabbit protein also) are derived from a precursor (propeptide), which contains 76 (human) or 79 (mouse) additional amino acids attached to the N-terminus of the mature hormone. Both the mature peptide and the precursor sequences are highly conserved between mouse and human; 79% of the residues in the mature hormone, and 86% of the residues in the propeptide sequence are conserved between the human and mouse forms^{48–50}. The conservation of the propeptide sequence has been taken to suggest^{48,49} that it fulfills a distinct biological function. Preliminary studies indicate that the 79-residue leader sequence is cut at no fewer than eight sites during protein processing, leaving a 2.9K peptide as the largest potentially intact segment (Fig. 2).

Cachectin is a moderately hydrophobic protein containing two cysteine residues per subunit which are believed to form a single intrachain disulphide bridge⁴². Hydrophobicity plots reveal seven regularly spaced hydrophobic peaks, suggesting the existence of uniform loops within the polypeptide chain that pass through a hydrophobic core. Electrophoresis of purified cachectin in SDS-containing gels under non-reducing conditions

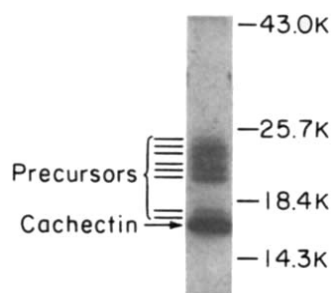


Fig. 2 Immunoblot of mouse cachectin from RAW 264.7 cells. Antisera raised against homogeneous solutions of the mature mouse hormone recognize this antigen, as well as eight precursor species on Western blots of secretory proteins derived from RAW 264.7 cells. Thus, during processing the prohormone is cut at multiple sites. When medium obtained from mouse peritoneal macrophage cultures is electrophoresed and blotted (not shown), a similar pattern is observed; however, cleavage of the prohormone is more complete. Human monocytes and peritoneal macrophages secrete neither the prohormone nor intermediate cleavage products in detectable quantities.

yields a single, 17K band, suggesting that no interchain bonds are involved in the formation of multimers.

Production, distribution and clearance

Cachectin is secreted by macrophages and travels via the circulation to distant sites in the body, where it binds to specific receptors^{13,53}, exerting discrete metabolic effects. The high levels of cachectin that circulate *in vivo* after challenge with endotoxin^{51,53} have facilitated detection of the hormone by means of the two bioassays noted above—in *vitro* tumour necrosis and lipoprotein lipase suppression.

Once induced by endotoxin, macrophages produce copious quantities of cachectin, amounting to approximately 1% of the total secretory product of the cells. In rabbits, a rise in plasma cachectin bioactivity occurs within minutes of intravenous (i.v.) administration of endotoxin⁵³, and nanomolar plasma concentrations of the hormone are achieved, corresponding to the production of milligram quantities of cachectin^{51,53}. The hormone is cleared from the plasma with a half-life of ~6 min⁵³. The major tissue targets of cachectin include the liver, skin, kidneys, lungs and gastrointestinal tract⁵³.

The plasma membrane receptor to which cachectin binds has not yet been isolated, and the nature of the signal transduced by binding remains unknown. The K_d of the hormone–receptor interactions has been estimated at 10^9 (refs 13, 54, 55) and 10^{12} (ref. 56) by different groups of investigators. It has been determined that several cell types carry between 1,000 and 10,000 copies of the receptor. Recent crosslinking studies have revealed the existence of two polypeptide chains (75K and 95K) that may participate in binding⁵⁶.

Gene expression

The cachectin (TNF) gene resides on chromosome 6 in man, and is (TNF) closely linked to the gene specifying lymphotoxin⁵⁷, a protein with approximately 30% sequence homology to cachectin^{11,15,57}. Presumably, both genes arose from a common ancestral sequence through a tandem duplication event. Both genes are in turn linked to the major histocompatibility complex, although the exact degree of proximity is unclear⁵⁷. The human cachectin (TNF) gene, like the lymphotoxin gene, contains three introns, one of which interrupts the sequence encoding the mature hormone.

Considerable homology has been noted in the 5'-untranslated region of the lymphotoxin and cachectin (TNF) genes⁵⁷, suggesting the possibility of coordinate control of gene expression. It has also been noted⁵⁰ that the mouse and human genes display, in the 3'-untranslated region, a conserved sequence composed entirely of A and T residues which is 33 nucleotides long. This

sequence is composed of tandem and overlapping repeats of an octameric sequence: TTATTTAT. This same sequence is conserved in the 3'-untranslated region of several other inflammatory mediators, including lymphotoxin, IL-1, granulocyte-macrophage colony-stimulating factor, and the interferons. Its precise function is unclear.

Isolated peritoneal macrophages contain detectable quantities of cachectin messenger RNA prior to LPS induction; however, no synthesis of cachectin is detectable in such cells^{58,59}. Production of cachectin mRNA is greatly increased on exposure of the cells to LPS. In addition, the nascent mRNA, as well as the cryptic message, is mobilized for translation in the presence of LPS^{58,59}.

The macrophages of endotoxin-resistant (C3H/HeJ) mice do not produce cachectin in response to endotoxin. In these animals, a mutation (*lps*^d) on chromosome 4 (ref. 60) decreases transcription of the LPS-induced cachectin gene and completely inhibits the mobilization of cachectin mRNA^{58,59}.

Similarly, glucocorticoids (for example, dexamethasone) in nanomolar concentrations strongly inhibit cachectin gene transcription and the mobilization of the cachectin mRNA once it has been formed, effectively preventing synthesis of the hormone. This phenomenon may largely account for the protective effect of glucocorticoid hormones administered to animals immediately before injection of LPS^{58,59}. It is interesting to note that, once translation of the cachectin mRNA is in progress, dexamethasone is not able to arrest cachectin production and secretion⁵⁹. This result offers a new explanation for the clinical observation that glucocorticoids are relatively ineffective in the treatment of shock if administered after sepsis has become established.

The post-transcriptional blockade imparted by the *lps*^d mutation can be circumvented by pretreatment of C3H/HeJ macrophages with γ -interferon. Through an unknown mechanism, γ -interferon permits the translation of cachectin mRNA in these cells, and normal amounts of cachectin can be produced *in vitro*, provided high concentrations of endotoxin are applied to the macrophages⁶¹.

An augmented response to endotoxin administration (and increased sensitivity to the lethal effect of endotoxin)³⁴ is observed in animals presensitized by infection with BCG²⁵, *Corynebacterium parvum*²⁶, *Propionibacterium acnes*^{3,33}, or other bacterial agents that stimulate reticuloendothelial hyperplasia (the 'priming' phenomenon). The mechanism of priming is incompletely understood; however, priming appears to increase the number of cachectin-producing macrophages present at the time of endotoxin injection, and to augment the transcriptional and translational response of these cells (with respect to cachectin biosynthesis). Also, it is possible that priming leads to the formation of a larger 'stored' pool of cachectin mRNA, which is mobilized for immediate translation in the presence of endotoxin.

Pathophysiological effects

A considerable body of evidence has implicated cachectin as a central mediator of the wasting that accompanies chronic invasive disease states. When administered to animals in moderate doses, cachectin induces a state of anorexia and ensuing weight loss, coupled with an apathetic, unkempt appearance reminiscent of cachexia observed under natural circumstances. Cachectin also seems to have a major role in the pathogenesis of Gram-negative (endotoxin-induced) shock. Indeed, it is possible that this single protein is responsible for most of the deleterious effects of endotoxaemia (for example, fever, metabolic acidosis, diarrhoea, hypotension, disseminated intravascular coagulation, and their consequences) which may lead to death. Not only does the administration of large doses of cachectin directly mimic the clinical syndrome produced by endotoxaemia (B.B., unpublished data), but passive immunization against cachectin substantially mitigates the lethal effect of endotoxin¹⁶.

Considerable variation is observed among different species of mammals both in sensitivity to cachectin administered and in the amount of cachectin produced endogenously after challenge with LPS. In mice (a relatively endotoxin-resistant species), daily i.v. administration of 10 µg of cachectin will usually cause death within 4–5 days (B.B., unpublished data). Larger doses (25–50 µg per animal) are often fatal within hours. In some species that are more sensitive to the biological effects of endotoxin, smaller doses of cachectin have adverse effects.

Several biological effects previously ascribed to another monokine, IL-1, are now known to be elicited by cachectin also. Cachectin is a potent pyrogen. It causes fever through a direct effect on the hypothalamic thermoregulatory centres, and through induction of IL-1 biosynthesis; hence, a biphasic fever curve is observed in rabbits following i.v. administration of cachectin⁶².

Recent studies have shown that enhanced production of prostaglandin E₂ and collagenase by human synovial cells and dermal fibroblasts is also induced by cachectin⁶³. This effect of the hormone, which may have a significant role in the pathogenesis of inflammatory diseases of joints and other tissues⁶⁴, was also previously ascribed to IL-1. Similarly, bone resorption⁶⁵ is stimulated *in vitro* by moderate concentrations of cachectin. It has been suggested that cachectin may, in fact, serve as an osteoclast activating factor. Cachectin can stimulate the production of a procoagulant activity by vascular endothelial cells⁶⁶, and can inhibit the expression of thrombomodulin on endothelial cell surfaces, which may relate to the clotting abnormalities associated with endotoxaemia. The same effect may be important in eliciting the haemorrhagic necrosis of tumours *in vivo*. Cachectin (TNF) activates neutrophil adherence and degranulation⁶⁷ as well as phagocytosis⁶⁸, and activates eosinophils⁶⁹, enabling them to destroy schistosomula *in vitro*. Given that these numerous bioactivities have been discovered within the past few months, it seems certain that many other effects will be identified.

Post-receptor effects

Cachectin suppresses the expression of specific mRNA molecules in target cell populations, leading to alterations in metabolic activity. In rat adipocytes⁷⁰, cachectin has been shown to suppress the biosynthesis of the glycerol-3-phosphate dehydrogenase mRNA, as well as several other differentiation-specific adipose mRNAs, without affecting the levels of 'household' mRNAs (for example, actin mRNA).

It has been noted that cachectin and lymphotoxin share a common receptor⁷¹. As such, many (and probably most) of the effects ascribed to cachectin may also be produced by lymphotoxin. Another early event, observed after the application of lymphotoxin to cells⁷², is the fragmentation of cellular DNA. This process is probably relevant to the *in vitro* cytolytic activity of both cachectin and lymphotoxin.

Production in diseases

Regional or systemic production of cachectin may be responsible for the pathological effects of a wide variety of inflammatory and neoplastic diseases. Inflammatory diseases of skin, gastrointestinal tract, joints, muscle and the central nervous system all involve characteristic invasion of tissues by mononuclear cells. It is likely that stimuli other than endotoxin are capable of causing cachectin production *in vivo*, and that cachectin is elaborated under such conditions. In view of the fact that transformed tissues often retain (or acquire) the ability to produce and secrete polypeptide hormones, certain tumours may produce cachectin autonomously, while others may stimulate mononuclear phagocytes to do so. The development of a ready method for the detection of cachectin production *in situ* will be of considerable benefit in the identification of disease processes in which the hormone has a significant role.

The impact of the catabolic effects of cachectin on the organism as a whole may be sufficient to account for the wasting that

accompanies many chronic infectious and neoplastic diseases. The anorexia observed in laboratory animals treated with cachectin²⁰, the pyrogenic effect of cachectin⁶², and its catabolic effects on energy storage tissues^{46,70,73} all suggest that the hormone orchestrates the complex metabolic changes that lead to cachexia.

Evolutionary function

The cachectin (TNF) gene is one of two modern genes derived from an ancient duplication event. The fact that only 30% homology remains between the lymphotoxin and cachectin (TNF) genes suggests that the two diverged several hundred million years ago⁷⁴. Thus, the ancestral gene may have pre-dated vertebrate evolution.

The predominant physiological function of cachectin remains to be determined. Presumably, the ability of cachectin to activate neutrophils and eosinophils, and to mobilize energy stores, confers an advantage on the host. However, when cachectin is released in large quantities, the resulting metabolic derangements can be catastrophic, leading ultimately to death of the individual. It is possible that the beneficial effects conferred by low levels of cachectin have been primarily responsible for the conservation of the cachectin gene, while the ill-effects imparted by high levels of the protein have allowed the removal of severely infected individuals from the population, thus preventing further disease transmission.

Alternatively, it is possible that cachectin has an indispensable function that is entirely distinct from any yet defined: for example, the hormone may be essential to normal developmental processes and has therefore been preserved despite the risks attendant on its release during infection. Further work will undoubtedly shed light on those functions that have ensured conservation of the cachectin gene.

Therapeutic possibilities

Considerable effort has been devoted to the large-scale production of cachectin (TNF) for use in trials as an antineoplastic agent. To date, no consensus has emerged as to the efficacy of the hormone. The initial isolation of TNF, and its use in this manner, was predicated on the assumption that a means of retaining the tumour necrotic effect of endotoxin, while dispensing with its toxic properties, had been found. It now appears, however, that cachectin (TNF) is the principal mediator of both endotoxin-induced shock and tumour necrosis.

Efforts aimed at modifying the cachectin molecule to produce a second-generation mediator capable of lysing tumours without eliciting fever, hypotension and coagulopathy are in progress. It is hoped that certain tumours will prove uniquely sensitive to the effects of cachectin, and may be destroyed by concentrations of the hormone that are tolerated by the affected individual.

On the other hand, antagonism of the effects of cachectin in an effort to mitigate shock, cachexia or various local inflammatory processes, may yield important clinical benefits. Glucocorticoid hormones, which are dramatically effective in preventing shock and treating inflammation, appear to suppress cachectin production completely, if administered before the mobilization of cachectin mRNA⁵⁹. However, immunosuppression, osteoporosis, hyperglycaemia, and many other effects of glucocorticoid therapy, limit its usefulness. Inhibition of cachectin bioactivity, at or beyond the level of the receptor, might prove to be a far more specific mode of therapy. Studies aimed at characterizing the cachectin receptor are in progress, and may contribute to the development of specific blocking agents. It has been shown that passive immunization of animals against cachectin can prevent endotoxin-induced death¹⁶. It remains to be seen whether immunization (either active or passive) might similarly protect against other deleterious effects of the hormone, and whether this procedure might be used to clinical advantage.

This work was supported by NIH grants AM01314 and AI21359, and a grant from the Rockefeller Foundation (RF no. 85078).

1. Beisel, W. R. *A. Rev. Med.* **26**, 9–20 (1975); *Am. J. clin. Nutr.* **30**, 1236–1247 (1977).
2. Filkins, J. P. *J. reticulendothel. Soc.* **25**, 591–595 (1979); *Pathophysiology of the Reticuloendothelial System*, 93 (Raven, New York, 1981); *J. reticulendothel. Soc.* **27**, 507–511 (1980).
3. Baracos, V., Rodemann, H. P., Dinarello, C. A. & Goldberg, A. L. *New Engl. J. Med.* **308**, 553–558 (1983).
4. Filkins, J. P. *Fedn Proc.* **44**, 300–304 (1985).
5. Dinarello, C. A. *Rev. infect. Dis.* **6**, 51–95 (1984).
6. Auron, P. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7907–7911 (1984).
7. Lomedico, P. T. *et al. Nature* **312**, 458–462 (1984).
8. Evans, C. H. *Cancer Immun. Immunother.* **12**, 181–190 (1982).
9. Aggarwal, B. B., Moffat, B. & Harkins, R. N. *J. biol. Chem.* **259**, 686–691 (1984).
10. Aggarwal, B. B., Henzel, W. J., Moffat, B., Kohr, W. J. & Harkins, R. N. *J. biol. Chem.* **260**, 2334–2344 (1985).
11. Gray, P. W. *et al. Nature* **312**, 721–724 (1984).
12. Nathan, C. F., Murray, H. W., Wiebe, M. E. & Rubin, B. Y. *J. exp. Med.* **158**, 670–689 (1983).
13. Beutler, B., Mahoney, J., Le Trang, N., Pekala, P. & Cerami, A. *J. exp. Med.* **161**, 984–995 (1985).
14. Beutler, B. *et al. Nature* **316**, 552–554 (1985).
15. Pennica, D. *et al. Nature* **312**, 724–729 (1984).
16. Beutler, B., Milsark, I. W. & Cerami, A. *Science* **229**, 869–871 (1985).
17. Kawakami, M. & Cerami, A. *J. exp. Med.* **154**, 631–639 (1981).
18. Kawakami, M., Pekala, P. H., Lane, M. D. & Cerami, A. *Proc. natn. Acad. Sci. U.S.A.* **912**–916 (1982).
19. Hotez, P. J., Le Trang, N., Fairlamb, A. H. & Cerami, A. *Parasite Immun.* **6**, 203–209 (1984).
20. Cerami, A., Ikeda, Y., Le Trang, N., Hotez, P. J. & Beutler, B. *Immun. Lett.* **11**, 173–177 (1985).
21. Coley, W. B. *Am. J. med. Sci.* **105**, 487–511 (1893); *Trans. Am. surg. Ass.* **12**, 183–212 (1894); *Am. J. med. Sci.* **112**, 251–281 (1896); *Bull. Johns Hopkins Hosp.* **65**, 157–162 (1896); *Am. J. med. Sci.* **131**, 375–430 (1906).
22. Nauts, H. C. *Cancer Research Institute Monograph No. 8*, 2nd edn (Cancer Research Institute, Inc., New York, 1980).
23. Shear, M. J., Turner, F. C., Perrault, A. & Shovelton, J. *J. natn. Cancer Inst.* **4**, 81–97 (1943).
24. Hartwell, J. L., Shear, M. J. & Adams, J. R. Jr. *J. natn. Cancer Inst.* **4**, 107–122 (1943).
25. Carswell, E. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **72**, 3666–3670 (1975).
26. Green, S. *et al. J. natn. Cancer Inst.* **59**, 1519–1522 (1977).
27. Matthews, N. *Immunology* **44**, 135–142 (1981).
28. Matthews, N. & Watkins, J. F. *Br. J. Cancer* **38**, 302–309 (1978).
29. Matthews, N. *Br. J. Cancer* **38**, 310 (1978); *Br. J. Cancer* **44**, 418–424 (1981).
30. Fisch, H. & Gifford, G. E. *Int. J. Cancer* **32**, 105–112 (1983).
31. Watanabe, N., Sone, H., Neda, H., Niitsu, Y. & Urushizaki, I. *Gan. No Rinsho* **30**, 1290–1292 (1984).
32. Haranaka, K., Satomi, N., Sakurai, A. & Haranaka, R. *Cancer Immun. Immunother.* **18**, 87–90 (1984).
33. Urushizaki, I., Niitsu, Y. & Watanabe, N. *Gan To Kagaku Ryoho* **11**, 1356–1368 (1984).
34. Ha, D. K., Gardner, I. D. & Lawton, J. W. *Parasite Immun.* **5**, 513–526 (1983).
35. Wood, P. R. & Clark, I. A. *Parasite Immun.* **6**, 309–317 (1984).
36. Satomi, N., Haranaka, K. & Kunii, O. *Jap. J. exp. Med.* **51**, 317–322 (1981).
37. Watanabe, N., Sone, H., Neda, H., Niitsu, Y. & Urushizaki, I. *Gan To Kagaku Ryoho* **11**, 1284–1289 (1984).
38. Mannel, D. N., Moore, R. N. & Mergenhagen, S. E. *Infect. Immunity* **30**, 523–530 (1980).
39. Zacharchuk, C. M., Drysdale, B.-E., Mayer, M. M. & Shin, H. S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6341–6345 (1983).
40. Helson, L., Green, S., Carswell, E. & Old, L. J. *Nature* **258**, 731–732 (1975).
41. Sugarman, B. J. *et al. Science* **230**, 943–945 (1985).
42. Aggarwal, B. B. *et al. J. biol. Chem.* **260**, 2345–2354 (1985).
43. Shirai, T., Yamaguchi, H., Ito, H., Todd, C. W. & Wallace, R. B. *Nature* **313**, 803–806 (1985).
44. Wang, A. M. *et al. Science* **228**, 149–154 (1985).
45. Rouzer, C. A. & Cerami, A. *Molec. Biochem. Parasit.* **2**, 31–38 (1980).
46. Pekala, P. H. *et al. Trans. Ass. Am. Physns* **97**, 251–259 (1984).
47. Mahoney, J. R. *et al. J. Immun.* **134**, 1673–1675 (1985).
48. Fransen, L. *et al. Nucleic Acids Res.* **13**, 4417–4429 (1985).
49. Pennica, D., Hayflick, J. S., Bringman, T. S., Palladino, M. A. & Goeddel, D. V. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6060–6064 (1985).
50. Caput, D., Beutler, B., Hartog, K., Brown-Shimer, S. & Cerami, A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
51. Abe, S., Gatanaga, T., Yamazaki, M., Soma, G. & Mizuno, D. *FEBS Lett.* **180**, 203–206 (1985).
52. Itoh, H. *Eur. Pat. Appl.* no. 84105149.3.
53. Beutler, B., Milsark, I. W. & Cerami, A. *J. Immun.* **135**, 3972–3977 (1985).
54. Baglioni, C., McCandless, S., Tavernier, J. & Fiers, W. *J. biol. Chem.* **260**, 13395–13397 (1985).
55. Tsujimoto, M., Yip, Y. K. & Vilcek, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7626–7630 (1985).
56. Kull, F. C., Jacobs, S. & Cuatrecasas, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5756–5760 (1985).
57. Nedwin, G. E. *et al. Nucleic Acids Res.* **13**, 6361–6373 (1985).
58. Beutler, B., Milsark, I. W., Krochin, N. & Cerami, A. *Blood* **66**, 83a (1985).
59. Beutler, B., Krochin, N., Milsark, I. W., Luedke, C. & Cerami, A. *Science* (in the press).
60. Watson, J., Kelly, K., Largen, M. & Taylor, B. A. *J. Immun.* **120**, 422 (1978).
61. Beutler, B. *et al. J. exp. Med.* (submitted).
62. Dinarello, C. A. *et al. J. exp. Med.* (in the press).
63. Dayer, J.-M., Beutler, B. & Cerami, A. *J. exp. Med.* **162**, 2163–2168 (1985).
64. Krane, S. M., Goldring, S. R. & Dayer, J.-M. *Lymphokines Vol. 7*, 75–136 (Academic, New York, 1982).
65. Bertolini, D. R., Nedwin, G., Bringman, T., Smith, D. & Mundy, G. R. *Nature* **319**, 516–518 (1986).
66. Stern, D. M. & Nawroth, P. P. *J. exp. Med.* **163**, 740–745 (1986).
67. Gamble, J. R., Harlan, J. M., Klebanoff, S. J., Lopez, A. F. & Vadas, M. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8667–8671 (1986).
68. Shalaby, M. R. *et al. J. Immun.* **135**, 2069–2073 (1985).
69. Silberstein, D. S. & David, J. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1055–1059 (1986).
70. Torti, F. M., Dieckmann, B., Beutler, B., Cerami, A. & Ringold, G. M. *Science* **229**, 867–869 (1985).
71. Aggarwal, B. B., Eessalu, T. E. & Hass, P. E. *Nature* **318**, 665–667 (1985).
72. Schmid, D. S., Tite, J. & Ruddle, N. H. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
73. Pekala, P., Kawakami, M., Vine, W., Lane, M. D. & Cerami, A. *J. exp. Med.* **157**, 1360–1365 (1983).
74. Kimura, M. *Nature* **217**, 624–626 (1968).

ARTICLES

Tropical convection, ionospheric potentials and global circuit variation

Ralph Markson

Center for Meteorology and Physical Oceanography, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

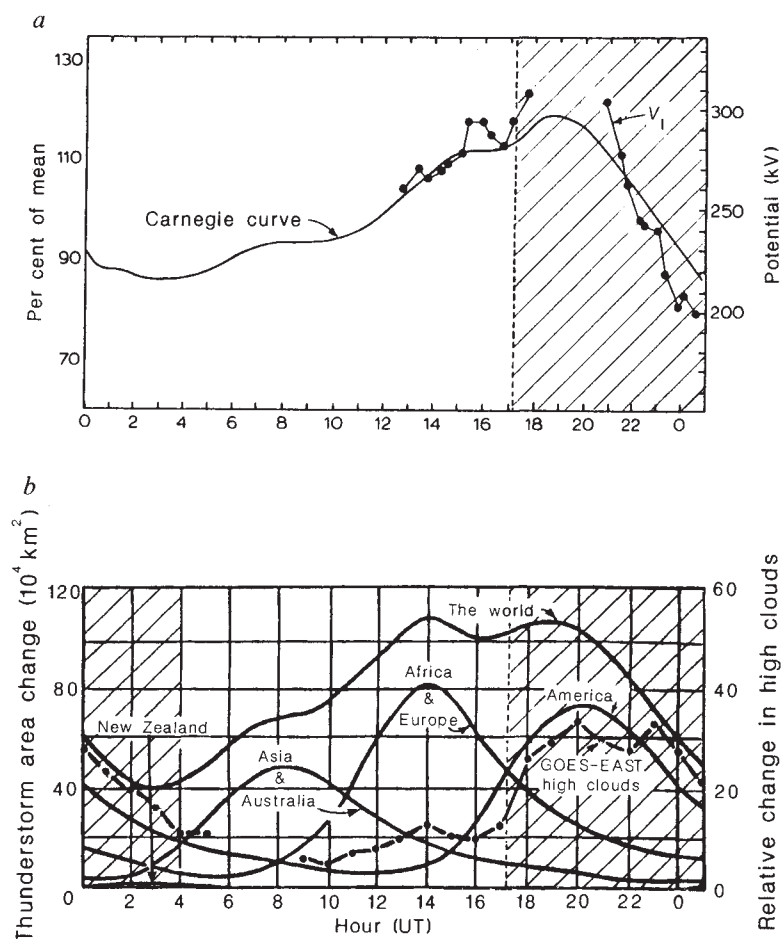
The ionospheric potential (V_I), a measure of the Earth's overall electric field intensity, correlates with the classic 'Carnegie curve' diurnal variation in electric field caused by the global distribution of thunderstorms. A comparison of V_I variation with satellite images of high tropical cloud fields indicates that equatorial thunderstorms over the continents are dominant in maintaining the Earth's electric field. V_I time series obtained at just one location may provide a high-resolution measure of intense tropical convection.

ONE of the most interesting properties of atmospheric electricity is that, unlike any other geophysical parameter, the Earth's fair-weather electric field intensity follows Universal Time (UT) rather than local (Sun) time. Its variation within each 24-h period is controlled by the integrated currents from globally distributed thunderstorms which are predominantly located within 20° of the equator in three regions: Africa, the Americas and the Maritime Continent (Indonesia, New Guinea and the western equatorial Pacific). The Scottish physicist C. T. R. Wilson was the first to explain the maintenance of the Earth's electric field^{1,2}. He recognized that thunderstorms cause the highly conductive Earth to acquire a negative charge, and that the positive charge flowing upward over thunderstorms is distributed rapidly and homogeneously throughout the highly conductive upper atmosphere (ionosphere) and returns to Earth in

non-thunderstorm regions through the slightly conductive atmosphere as a conduction current. Thus, the fair-weather electric field in all parts of the world is influenced equally and simultaneously by thunderstorm currents regardless of their location, and the ionosphere is at a constant potential throughout. Most researchers consider this model to be correct because it provides the only reasonable explanation for the atmospheric electric field: thunderstorm currents are the only significant means for charging the earth.

Today this spherical capacitor model remains essentially unchanged except that we have come to recognize that upper-atmosphere electrical generators can modulate atmospheric electric fields in the polar cap region (15% of the Earth's surface)^{3–6}. Measurements by the MIT Millstone Hill radar indicate that steady-state dawn-to-dusk potentials rarely exceed

Fig. 1 *a*, Correlation of aircraft-derived V_1 time-series obtained during two flights over the Atlantic Ocean near Boston on 7 September 1984 with the normalized Carnegie curve ($r = 0.97$, $n = 22$, $P < 0.0001$). The aircraft data were obtained by integrating the vertical electric field measured while ascending or descending between the ocean surface and a height of 6.6 km. The Carnegie curve is the average of electric field measurements made far from land during global magnetic surveys conducted by the *Carnegie* in the 1920s¹⁰. A clean, cloud-free airmass allowed these low-noise measurements which demonstrate the repeatability of the Carnegie curve in a single day's data. The period during which it is appropriate to compare V_1 with high-cloud variations (*b*) is cross-hatched. *b*, Comparison of the variation of highest clouds over South America, Central America and Mexico on 7 September 1984 (broken line) with the diurnal variation of the world's thunderstorm areas from Brooks' 1925 analysis³⁰. The cloud analysis was performed on the Airborne Research Associates pictorial image analysis system, using 16-mm cinefilm obtained from the GOES-EAST infrared scanner. Brooks' composite curve for the world is derived from the three sector curves (solid lines). A correspondence between ionospheric potential and the high clouds is expected only during the 17:00–04:00 UT interval (hatched), when the America sector is the major contributor to the global thunderstorm generator.



60 kV, although occasional transients of limited extent can exceed this value (ref. 5 and W. L. Oliver, personal communication). Such potentials are superposed on a potential averaging ~ 260 kV (ref. 6), maintained by the global thunderstorm generator. Because the ionosphere at sub-auroral latitudes can be considered as essentially an equipotential surface, it is possible with only one reliable measurement of V_1 at sub-auroral latitude to quantify the intensity of the d.c. global circuit current and the Earth's overall electric field intensity. This has been done by integrating potential gradient profiles obtained by aircraft^{6–8} and balloon⁹ soundings spanning 85–95% of the Earth-ionosphere potential difference, and extrapolating the remainder.

Historically, observing the variation of the global circuit current has been one of the fundamental problems in atmospheric electrical research. Ground-level or lower-atmosphere electric field measurements within the exchange layer are generally incapable of accurately detecting the UT diurnal variation of the fair-weather electric field in data from a single day because of perturbations stemming from both natural and man-made aerosols and space charge under the influence of atmospheric boundary layer processes. Even on a ship at sea it generally requires averaging at least a week's data to bring out the Carnegie curve UT diurnal variation¹⁰. We have avoided these perturbations by making aircraft V_1 soundings and constant-altitude electric field measurements in low-convection regions of clean, cloud-free air over the ocean, thus allowing the routine observation of diurnal variation in a single 24-h time-series^{6,7,11}. In addition to the ability to make such measurements at selected low-noise regions away from clouds, haze, mountains and convection, aircraft measurements allow the effects of cloud-scale variations in columnar resistance and/or space charge within the exchange layer to be averaged out by the vehicle's motion.

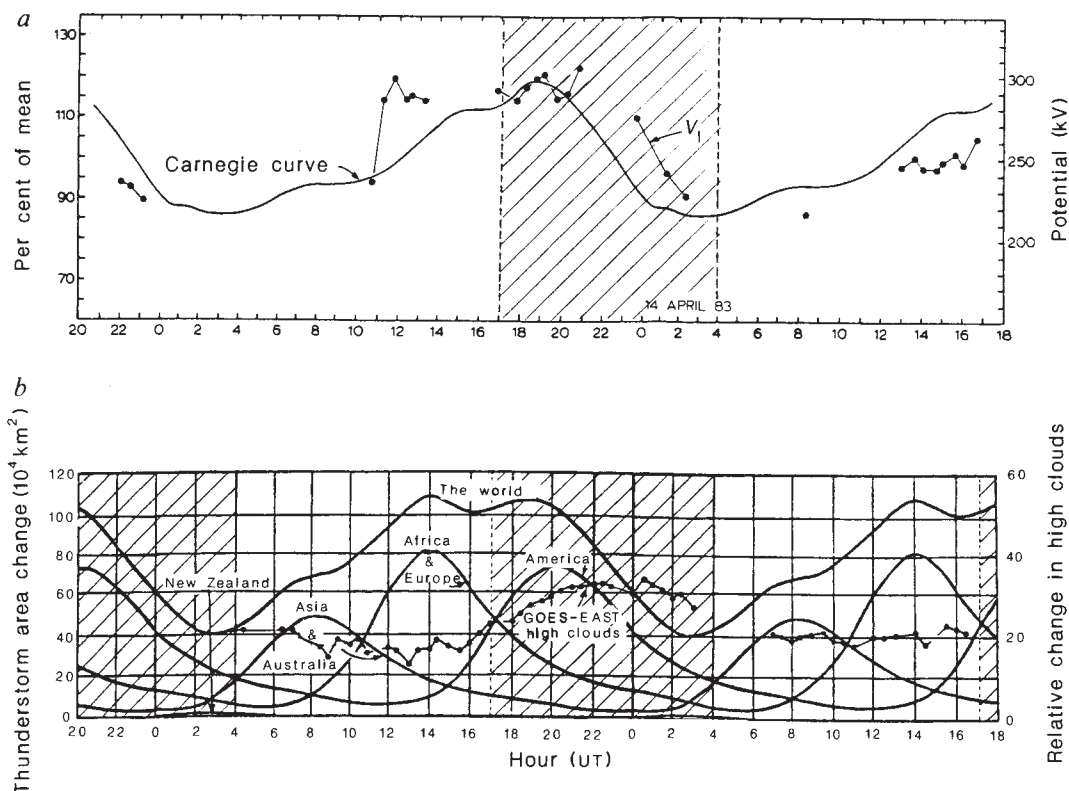
Simultaneous V_1 soundings are made at widely separated locations using two aircraft⁶: when such time-series match it can be assumed that local effects have not contaminated the data and that global-scale variations have been observed.

Carnegie curve UT variation and V_1 data

The Carnegie curve is the average variation of the Earth's fair-weather electric field intensity over a 24-h period. It was originally defined in analysis of data obtained during global magnetic survey cruises of the research vessel *Carnegie*¹², and has subsequently been replicated in averages of many other data sets¹³. This quasi-periodic diurnal variation (Fig. 1a) was found to follow Universal Time¹⁴ and to correlate with an early estimate of global thunderstorm variability¹⁵. Maxima in the three components making up the composite thunderstorm curve (Fig. 1b) occur when it is afternoon in Africa, America and Indonesia, and the minimum in this curve occurs when it is afternoon over the central Pacific Ocean—a region of few thunderstorms. The similarity between the Carnegie curve and the temporal variation of worldwide thunderstorm activity has been one of the cornerstones supporting the Wilson circuit concept.

With the ability to measure variations in the integrated output of global thunderstorms (inferred from V_1 time-series) and the recent availability of continental-scale thundercloud statistics (from satellite observations), it is now possible to investigate the Wilson circuit hypothesis through direct comparison of the Earth's electric field intensity and simultaneous thunderstorm variations. Aircraft measurements have been conducted in which a series of V_1 soundings extending over periods of 1/2 day and almost 2 days (with some breaks) were obtained. For the first time the magnitude of V_1 has been recorded with high temporal resolution (1/2 h). (Temporal resolution is limited by the time

Fig. 2 *a*, Correlation of variation in V_1 measured on a series of flights over the Atlantic Ocean off Virginia during the period 13–14 April 1983 with the Carnegie curve, as in Fig. 1*a* ($r = 0.54$, $n = 29$, $P < 0.01$). The aircraft soundings were made between the ocean surface and a height of 5.0 km. The period during which it is appropriate to compare V_1 with high-cloud variations (shown in *b*) is hatched. *b*, Comparison of high-cloud variations over South America on 13–14 April 1983 with Brooks' thunderstorm area curves as in Fig. 1*b*. Comparison between V_1 and high clouds is meaningful only in the 17:00–04:00 UT interval (hatched).



it takes to complete a sounding. Details regarding the sounding procedure and the estimated accuracy of $\pm 3\%$ are provided in refs 6, 7.) To perform this study, thundercloud variations have been analysed for two of the sampling periods and compared with the variation of ionospheric potential.

The first V_1 time-series was obtained during a sequence of flights over the Atlantic Ocean, 100 km off the coast of Wallops Island, Virginia, on 13–14 April 1983. These data correlated significantly with the Carnegie curve ($r = 0.54$, $n = 29$, $P < 0.01$), as seen in Fig. 2*a*. During this experiment it was necessary to work on a prearranged schedule, so the flights were made under hazy conditions which extended out to sea well beyond 100 km; the haze introduced some noise into the measurements. The most prominent deviation from the Carnegie curve occurs at 06:00 local time (LT) on 13 April. That this was the time of dawn may be significant: Muir¹⁶ has suggested that the atmospheric electrical 'sunrise effect'¹⁷ (an abrupt rise in field intensity at sunrise) may result from the first impingement of the Sun's rays on the upper atmosphere. Another source of error was that the soundings were made to a height of only 5.0 km to minimize the time between measurements.

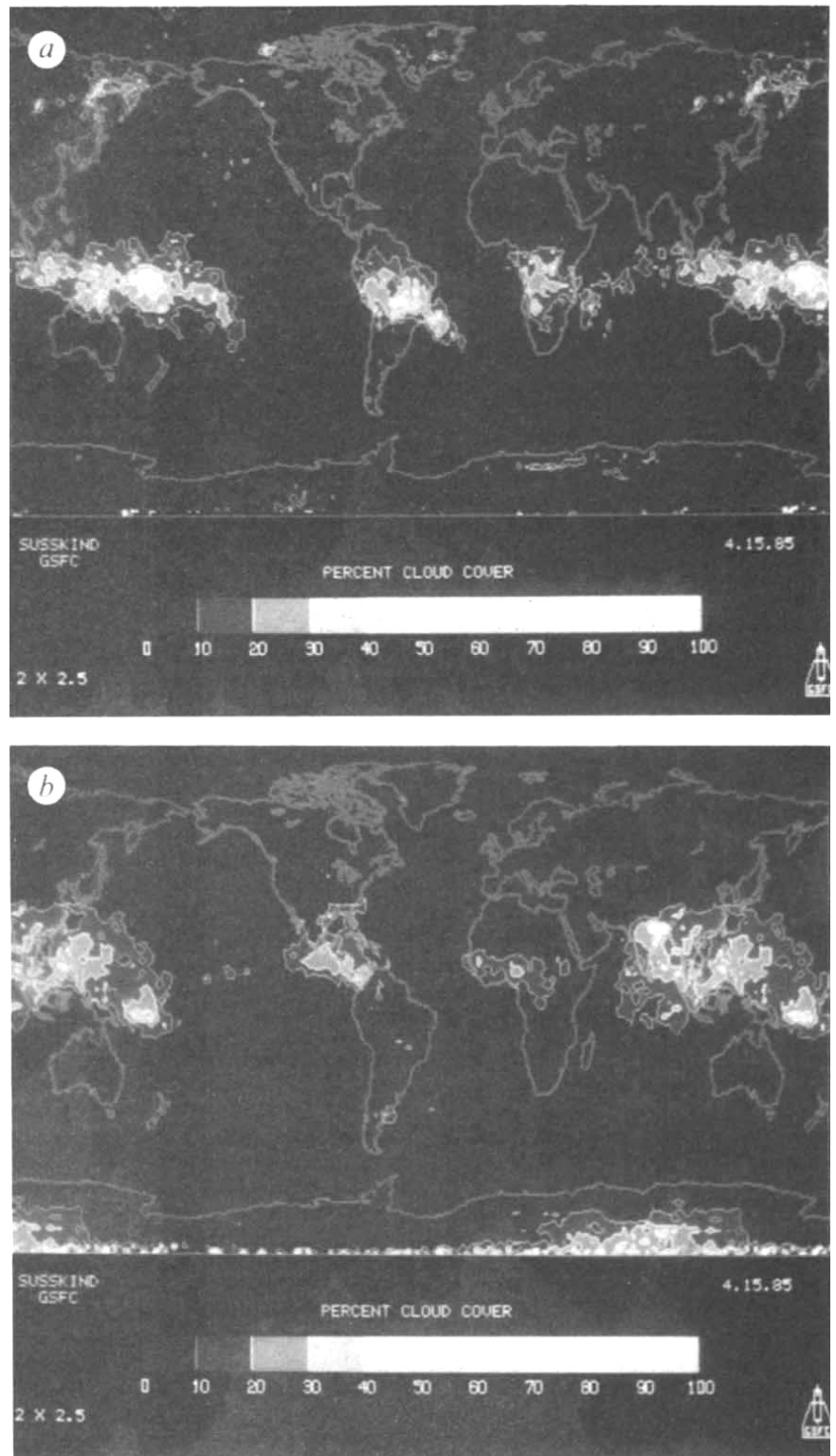
A second series of closely spaced V_1 soundings was obtained under favourable conditions in a clean, largely cloud-free air-mass over the ocean on 7 September 1984. The aircraft conducted 22 V_1 soundings to a height of 6.6 km over a 12-h period ~ 80 km north-east of Boston. (An electrical problem on the aircraft prevented extending this time-series.) The V_1 data and the Carnegie curve are highly correlated ($r = 0.97$, $n = 22$, $P < 0.0001$), as seen in Fig. 1*a*. There has been only one other sequence of V_1 soundings performed over an extended period (that of 23–24 April 1983), and this again correlates with the Carnegie curve ($r = 0.97$, $n = 6$, $P < 0.01$). (These observations, reported in ref. 6, have not been compared with satellite imagery.) Thus, during each of the three periods when there has been an opportunity to make a series of closely spaced V_1 measurements, these time-series have correlated with the Carnegie curve at the 99% confidence level or higher. Thus, all analyses of closely spaced V_1 time-series indicate that a daily variation similar to the Carnegie curve is to be expected in the

data from a single day, and that this variation can be considered as diagnostic of global circuit current variation (as opposed to meteorologically induced fluctuations).

Rapid variations?

Recently, Holzworth *et al.*¹⁸, reporting electric-field data obtained with constant-altitude stratospheric balloons, concluded that there is considerably greater variability of the global circuit current than previously realized. Their data indicated fluctuations on the order of 70% ('twice the amplitude of the average diurnal variation') over periods as short as tens of minutes. However, the Earth's electric field results from the integrated currents from $\sim 2,000$ – $3,000$ globally distributed thunderstorms¹⁹, and it is therefore statistically improbable that the charging current will undergo rapid large variations. The $1/e$ time constant for the atmospheric electric field (the time required for atmospheric space charge to be redistributed and the surface charge density of the Earth to reach new equilibrium) is controlled by the relatively low-conductivity air near the Earth's surface. Since this time constant in clean oceanic air at mid-latitude is ~ 10 min (ref. 6), it would require an abrupt 75% change in the global thunderstorm generator output to cause a 70% variation in the global circuit current in 30 min. Dolezalek²⁰ discusses how constant-altitude aircraft or balloon measurements at any height are affected by changes in columnar resistance caused by clouds or haze in the lower atmosphere. Such conductivity variations modulate the air–Earth current density and thus the electric field intensity throughout the vertical column from the Earth's surface to the ionosphere. Charged clouds (including those not producing lightning) could also affect stratospheric electric fields. (See ref. 21 for a discussion of the effects of lower-atmosphere processes on atmospheric potentials.) Analysis of the 1/2-h-resolution aircraft-derived V_1 time-series indicates that the greatest rate of change in Fig. 1*a* (from 23:02 to 23:32 UT) is 13% per 30 min and that in Fig. 2*a* (from 10:41 to 11:15 UT, during the anomalous dawn period) is 19% per 30 min. Since these are selected 'worst cases' based on only a few data points and there is some error in the measurement, it is probable that they are overestimates. In Fig. 1*a* the

Fig. 3 Spatial distribution of high clouds above 12 km (<200 mbar) over the Earth's surface for *a*, January and *b*, July 1979. Electric currents from thunderstorms in three global activity centres near the equator over South and Central America, Africa and the Maritime Continent control the temporal variation of the Earth's atmospheric electric field intensity. Analysis provided by Joel Susskind, NASA Goddard Space Flight Center, using HIRS2/MSU radiance data from the TIROS-N satellite²².



relatively rapid rate of decrease between 20:45 and 22:15 UT constitutes only a 6.5% decrease in V_1 per 30 min. Based on the three recent aircraft-derived V_1 time-series it appears that the variation in the Earth's overall electric field intensity is generally less than 10%, and rarely exceeds 15%, per 30 min.

Satellite cloud imagery

Although ionospheric potential can provide a measure of the integrated currents from global thunderstorms, it is also desirable to examine the spatial and temporal variation of the thun-

derstorms themselves for comparison with V_1 variation and the Carnegie curve. The advent of satellites, with their global perspective and infrared sensors capable of measuring cloud-top temperatures day and night, makes such a test of the Wilson circuit hypothesis possible for the first time. The geostationary satellites located at fixed positions over the equator, with a view covering 120° of longitude, are particularly well placed to obtain such data; they provide an excellent perspective from which to observe the spatial and temporal changes of the deep convective clouds in the equatorial-tropical latitude belt which contains

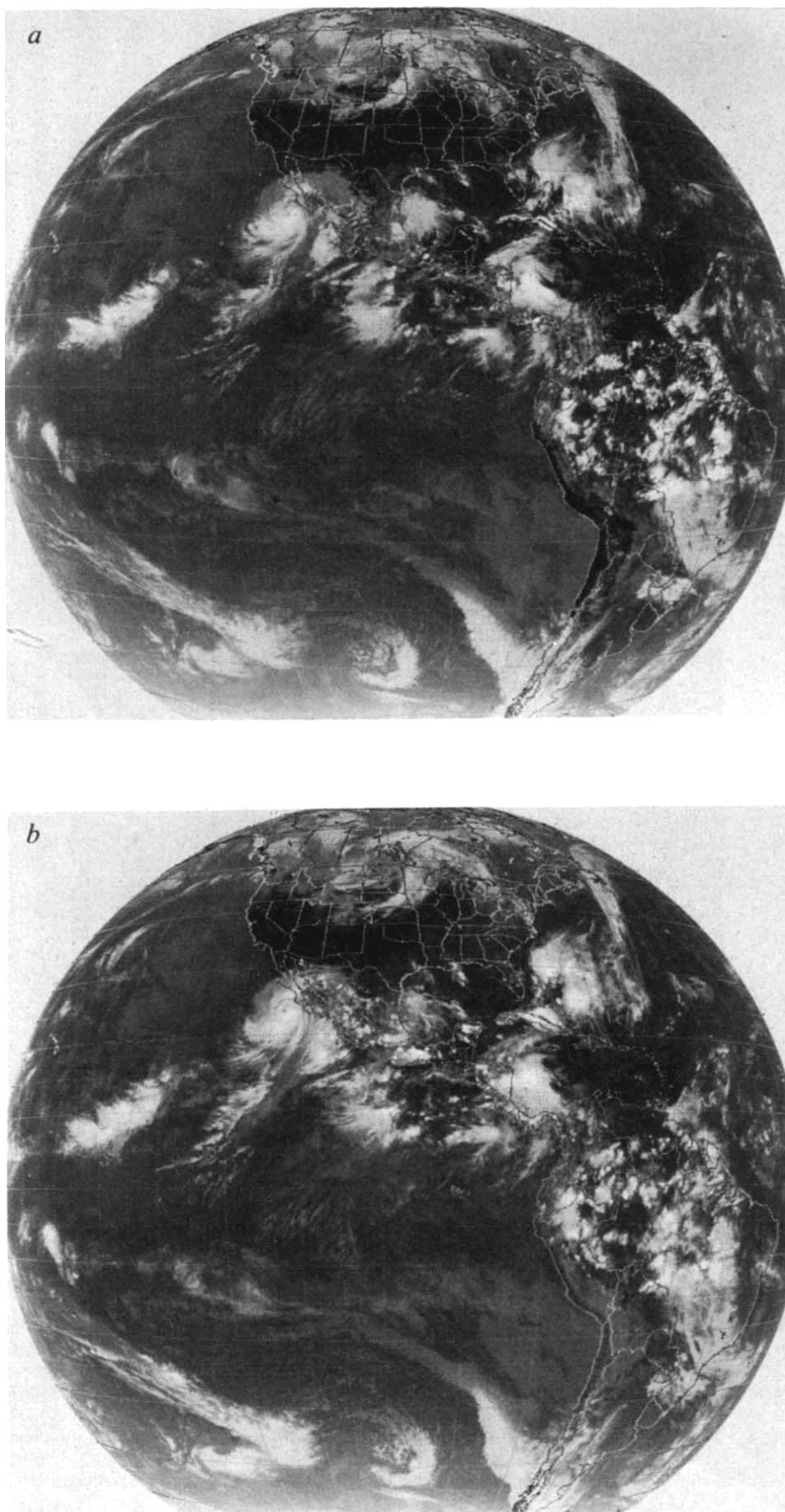


Fig. 4 Infrared cloud images from GOES-EAST. *a*, At 20:00 UT (16:00 LT in the Amazon basin) on 7 September 1984, the time of the first maximum in the high-cloud area curve seen in Fig. 1*a*. Note the many cellular, deep convective cloud complexes over South America and the lack of similar structures over Mexico. A hurricane is off Baja California. *b*, At 23:00 UT on 7 September 1984, the time of the second maximum in the high-cloud variation. It is now 16:00 LT in Mexico, where convective cells (thunderstorms) have developed. Close inspection indicates somewhat fewer cells and a softening in appearance of the deep convective regions over South America, and an increase in cirrus residue; for example, in Colombia, the Amazon basin and just north of Panama.

most of the global thunderstorm generator. The Japanese GMS satellite over Indonesia, the US GOES-EAST over the eastern Pacific and South America and the European METEOSAT over western Africa cover the three centres of maximum global thunderstorm activity. These three regions are seen near the equator in Fig. 3a and b, which depicts cloudiness above the 200-mbar pressure level (~ 12.5 km) in January and July 1979²². The 200-mbar level was the highest accessible to analysis by the current program, which uses $125 \text{ km} \times 125 \text{ km}$ box averages. This misses the smaller, deepest convective structures which exceed the 100-mbar level. Note that there is greater high cloudiness over South America and Africa in the Northern Hemisphere winter, whereas the high cloud cover over the Maritime Continent peaks in the Northern Hemisphere summer. The average latitude of the centre of high cloudiness moves from $\sim 10^\circ \text{ N}$ to 10° S with the seasons, in agreement with past studies of the annual migration of the Intertropical Convergence Zone (ITCZ) position. The global cloud analysis shows that the area of high clouds over the Maritime Continent is larger than over the other two thunderstorm centres. However, as the V_I measurements indicate that the largest daily values occur when thunderstorm activity peaks in the afternoon over South America and Africa, one may infer that there are more thunderstorms per unit area over South America and Africa than over the Maritime Continent (mostly water), where a smaller percentage of the high clouds are associated with energetic convection and cloud electrification. In an earlier study using ground-station observations, Fischer²³ found that while thunderstorm activity at mid-latitude peaks in northern summer, within 20° of the equator (and particularly within 10°) it peaks in northern winter.

Why does the field peak in winter?

The satellite data help to explain the apparent paradox that fair-weather electric field intensity^{24,25} and V_I ²³ are greater in the Northern Hemisphere winter than in the summer (by about 30%, from the *Carnegie* data¹²) despite the fact that thunderstorm activity over land is a summer phenomenon at mid-latitude and in the subtropics, and most of the Earth's land mass is in the Northern Hemisphere. Several factors explain the apparent importance of equatorial convection over land in maintaining the Earth's electric field. The tropopause, which cuts off convective cloud development, is at ~ 13 km altitude at mid-latitude and ~ 18 km at the equator; because the current density of a thunderstorm is proportional to the cube of its height^{26,27}, the equatorial storms would produce a current 2.65 times as large as that from mid-latitude storms per unit cross-sectional area of cloud. In addition, the equatorial storms are wider as well as higher, so the generator area is larger. Also, the current from equatorial storms would couple more efficiently to the global circuit because most of the electric field lines over the higher tropical convection would extend up to the ionosphere, whereas over the lower clouds more of the lines would bend back to Earth near the storms. Stronger solar heating of the Earth's surface and more atmospheric water vapour are also important factors favouring equatorial convection over land. Finally, the thunderstorm-producing area within 20° of the equator is several times larger than combined thunderstorm regions over the rest of the globe. Inspection of satellite imagery leaves little doubt that the global thunderstorm generator is located at low latitudes in all seasons. This is illustrated in the January and July cloud analyses (Fig. 3a, b) and in the GOES satellite imagery (Fig. 4). In the infrared imagery (Fig. 4), the highest (coldest) clouds are the brightest, and deep convection can be identified by its cellular structure.

To quantify the variation in thunderstorm activity, we have developed a satellite image analysis system capable of digitizing and analysing infrared cloud cinefilms containing every satellite image. The infrared imagery provides a coherent data set, day and night, with 1/2-h time resolution for the GOES data and 3-h resolution for the GMS, and the analysis accepts only those

signals with brightness (temperature) levels above (below) a given threshold, so that deep convective clouds reaching the tropopause can be identified. It is assumed that there will be close correlation between the number of the thunderstorms and the area of high tropical and equatorial clouds over Africa, South America and land masses in the Maritime Continent. Over the ocean this assumption is not valid: the high convective clouds that generally exist in the ITCZ (within 15° of the equator) are rarely electrified sufficiently to produce lightning. Continental Airlines pilots flying routinely from Hawaii to Australia across the ITCZ report that lightning is uncommon in this region of consistent high cloudiness. Also, in a study of the languages of people living in the Pacific Ocean region, it was found that the words for thunder and lightning become less frequent for locations approaching the equator until, close to the equator, the words do not exist (R. Taylor, personal communication). Studies of global-scale nocturnal lightning distribution indicate a ratio of land to ocean flashes of about 10:1 when coastal thunderstorms are classified in the 'land' category^{28,29}, and when oceanic lightning is investigated it is frequently found to occur over an island (R. Orville, personal communication).

We have examined three years of GOES-EAST images covering the thunderstorm regions of Mexico and Central and South America, and two years of images from the Japanese GMS satellite. An increase of convective activity in the mid- to late afternoon (typically at $\sim 16:00$ LT) is observed every day of the year without exception in the equatorial-tropical latitude zone over the land masses and surrounding waters in both regions, which constitute two of the three activity centres of the global thunderstorm distribution. Although African thunderstorm images have not yet been examined, there is no reason to expect different conditions there. The remarkable regularity in timing and global position of the deep convective activity centres is consistent with the generally accepted conclusion that the global circuit current and its diurnal variation (similar to the Carnegie curve) are controlled by thunderstorm activity over the continents, which peaks in the mid- to late afternoon.

Comparison of V_I and high-cloud variations

Figure 1 illustrates the previously described correlation between the average of the *Carnegie* electric field measurements over the ocean and the variation of global thunderstorms based on the climatological analysis of Brooks³⁰, which was an estimate of global thunderstorm area variation based on ground observations. The composite curve of global thunderstorm temporal variation is derived from the 'Asia and Australia', 'Africa and Europe' and 'America' components; the maxima in these curves occur at 16:00 LT near the corresponding equatorial centres in Indonesia, Africa and South America. Also shown (arbitrary units) is the variation in area of highest clouds in the 'America' sector derived from our analysis of GOES-EAST infrared imagery on 7 September 1984 during the period when the V_I data were obtained near Boston. The high clouds over Mexico and Central and South America, including the ocean areas east and west of Mexico and Central America, and north of South America, were analysed; the remaining lower clouds did not exceed the brightness threshold. There is good agreement between the observed cloud variation and the 'America' component of thunderstorm variation. The persistence of high cloudiness following the first maximum at 20:00 UT (16:00 LT), due to thunderstorms over South America, is caused by two factors: (1) cirrus 'blow-off' outflow from the tops of thunderstorms which typically remains near the tropopause for several hours or longer following the convective development phase, and (2) the development of convective clouds over Mexico, 3 h west of central South America, leading to a second maximum at 23:00 UT (16:00 LT in Mexico). As thunderstorm electrification is closely tied to convective activity, it is expected that the thunderstorm currents maintaining the global circuit would be correlated with the development and mature stages

of thunderstorms rather than their dissipation phase. The cirrus clouds are left near the height of the cloud tops like foam on the ocean surface long after the breaking waves that caused it have dissipated. These observations are similar to those of Minnis and Harrison³¹, who found for the average diurnal variation of high clouds in the Amazon basin at 62° W a rapid rise in the afternoon, peaking near sunset (~18:00 UT), followed by a slow decrease until ~22:00 UT. Because the cloud analysis reported here covers a 3–4-h-wide latitude range crossing South and Central America, the period of maximum convective activity is more extended in time.

The infrared cloud images from GOES-EAST (Fig. 4a, b) are useful for interpreting Fig. 1. A comparison of these images, obtained at the time of the two peaks in high cloudiness, shows a decrease in number of the many relatively small cellular convection centres (active thunderstorm complexes), an increase in the area of more diffuse cirrus clouds (seen most easily over and off the coast of Ecuador, over northern Peru, and over Colombia and the Amazon basin), and the appearance of thunderstorms over Mexico during the later time period. During the hours from ~17:00 to 04:00 LT (hatched in Fig. 1b) the 'America' component is the major contributor to the composite ('The world') thunderstorm curve; thus, it is meaningful to compare the variation in deep convection in the 'America' sector with V_i only during this interval. The maximum V_i on this day occurs between 18:00 and 21:00 UT, at the time of the first cloud maximum associated with developing and mature thunderstorms over South America. Thunderstorms in the 'America' sector have proportionately less influence on the total global circuit current after the maximum phase of the 'America' component, when clouds in the Indonesia sector start to become more important. It is interesting to note the agreement between the variation of high clouds from the GOES-EAST image analysis and Brooks' 'America' curve derived 60 years ago. Because of several factors, including the persistence of high clouds, the imprecisely known relationship of convective activity to conduction current intensity, and small sample size, a correlation between V_i and high clouds has limited meaning. The maximum correlation coefficient for data in the 17:00–04:00 UT interval occurred when the clouds were lagged by 4 h ($r = 0.93$, $n = 8$).

Figure 2b shows a similar analysis for the period when the V_i measurements off Wallops Island were obtained on 13–14 April 1983. Only the cloud field over and adjacent to South America was analysed because there were no thunderstorms over Mexico and Central America. Again, in the 17:00–04:00 UT period there is correspondence between the high clouds and the 'America' component of thunderstorm variation, with the cloud maximum lagging by a few hours. The baseline value of the high-cloud curve during the 04:00–17:00 UT period appears higher, compared with Brooks' 'America' curve, than in Fig. 1 because characteristics of the cloud images made it necessary to use a warmer-temperature detection threshold than on the other day; thus, more of the less variable middle-level clouds were detected. The V_i time-series tends to follow the 'America' component thunderstorm statistics better than the high-cloud area in the 17:00–04:00 UT interval; as on the other

day, there is an apparent persistence of high cloud after the V_i maximum, presumably due to the residence time of the thundercloud-generated cirrus. As in the other case, inspection of the corresponding infrared pictures from the time of the largest V_i (at the peak of the 'America' curve) to the end of this continuous sequence of measurements (when V_i is at a minimum) shows a decrease in the ratio of cellular convection to diffuse cirrus residue, while the total area of high cloud remains relatively constant. It is difficult to identify active convective clouds within a veil of high clouds in satellite imagery. The correlation between V_i and high clouds in this case peaks when the clouds are lagged by 2 h ($r = 0.90$, $n = 9$).

The above direct comparisons of V_i time-series with thundercloud variations is the first attempt at this type of analysis. So far two intervals have been examined. Although study of such a limited database cannot provide conclusive results, the correlation of high cloud cover over Mexico, Central America and South America with V_i variation in a one-day's time-series suggests the potential for this approach in global-circuit research.

V_i as a proxy index of Hadley cell intensity

Research on the global circuit, which has sometimes been considered to be a curious and unimportant part of atmospheric science, may have unexpected practical application. As energetic convection is highly correlated with thundercloud electrification, P. H. Stone (personal communication) has suggested that V_i intensity may be a good indicator of tropical condensation and the Hadley (global convection) cell which it forces. The analyses and considerations above suggest that V_i intensity is in some ways a better indicator than high-cloud variations of vigorous tropical convection. Thus, in addition to being a measure of the intensity of the global circuit current, V_i may serve as a real-time proxy index for the globally integrated strength (poleward momentum transport) of the Hadley cell, a parameter which is currently known with a temporal resolution of only 1 month, from considerable analysis of data from the rawinsonde balloon network. Information on the longitudinal variation of convection could also be derived from the time of day, because the major contributions to V_i from each of the three continental regions occur during local afternoon. As Hadley cell circulation can affect mid-latitude circulation with time delays of a few days, such information could provide improved weather forecasting as well as testing of models.

I thank J. Hansman, T. Thompson, W. Hoffman, H. Schnerr, S. Bussolari, G. Pollett, C. Baker, L. Spaiser and B. Anderson for serving as pilots and observers in the aircraft programme. S. Dolan performed the satellite image analysis, B. Anderson analysed the aircraft data and assisted in preparation of this report and J. Susskind provided special analyses of high cloud distributions from his unpublished data. P. H. Stone first suggested the possible use of V_i as a proxy index for Hadley cell intensity. The ionospheric potential measurements were supported by NSF grant ATM-8211403 to MIT and NASA contract NAS6-3203 to Airborne Research Associates (ARA). The satellite image analysis was funded by NASA grant NRG22-009-727 to MIT and NASA contract NAS8-84771 to ARA.

Received 29 April 1985; accepted 11 February 1986.

1. Wilson, C. T. R. *Proc. Camb. phil. Soc.* **13**, 363–382 (1906).
2. Wilson, C. T. R. *Phil. Trans. R. Soc. A* **221**, 73–115 (1920).
3. Park, C. G. & Dejnakarindra. in *Electrical Processes in Atmospheres* (eds Dolezalek, H. & Reiter, R.) 536–543 (Steinkopff, Darmstadt, 1977).
4. Roble, R. G. & Hays, P. B. *J. geophys. Res.* **84**, 7247–7256 (1979).
5. Oliver, W. L., Holt, J. M., Wand, R. H. & Evans, J. V. *J. geophys. Res.* **88**, 41–54 (1980).
6. Markson, R. *J. geophys. Res.* **90**, 5967–5977 (1985).
7. Markson, R. *J. geophys. Res.* **81**, 1980–1990 (1976).
8. Clark, J. F. in *Recent Advances in Atmospheric Electricity* (ed. Smith, L. G.) 61–73 (Pergamon, New York, 1958).
9. Mühleisen, R. *Pure appl. Geophys.* **84**, 112–115 (1972).
10. Kasemir, H. W. *Pure appl. Geophys.* **100**, 75 (1972).
11. Anderson, R. V. *J. geophys. Res.* **74**, 1697–1700 (1969).
12. Parkinson, W. C. & Torrison, O. W. *UGGI (Sect. terr. Magn. Elect.) Bull. No. 8*, 340–345 (1931).
13. Israel, H. *Atmospheric Electricity II*, 350 (National Technical Information Service, Springfield, 1973).
14. Mauchly, S. J. *Terr. Magn. atmos. Elect.* **28**, 61–81 (1923).

15. Whipple, F. J. W. & Scrace, F. J. *Geophys. Mem., Lond.* **68**, 1–20 (1936).
16. Muir, M. S. *J. atmos. terr. Phys.* **37**, 553–559 (1975).
17. Israel, H. *Atmospheric Electricity II*, 384, 404 (National Technical Information Service, Springfield, 1973).
18. Holzworth, R. H., Onsager, T., Kinter, P. & Powell, S. *Phys. Rev. Lett.* **53**, 1398–1401 (1984).
19. Chalmers, J. A. *Atmospheric Electricity* 2nd Edn, 307 (Pergamon, New York, 1967).
20. Dolezalek, H. *Pure appl. Geophys.* **100**, 33–34 (1972).
21. Markson, R. *J. geophys. Res.* **89**, 2629–1635 (1984).
22. Susskind, J., Rosenfeld, J., Reuter, D. & Chahine, M. T. *J. geophys. Res.* **89**, 4677–4697 (1984).
23. Fischer, H. J. Univ. Tübingen (1962).
24. Chalmers, J. A. *Atmospheric Electricity* 2nd Edn, 169 (Pergamon, New York, 1967).
25. Israel, H. *Atmospheric Electricity II*, 353–354 (National Technical Information Service, 1973).
26. Vonnegut, B. *Met. Monogr.* **5**, 224–241 (1963).
27. Williams, E. J. *geophys. Res.* **90**, 6013–6025 (1985).
28. Orville, R. E. & Spencer, D. W. *Mon. Weath. Rev.* **107**, 934–943 (1979).
29. Orville, R. E. *Mon. Weath. Rev.* **109**, 391–395 (1981).
30. Brooks, C. E. P. *Geophys. Mem. Lond.* **24**, 147–164 (1925).
31. Minnis, P. & Harrison, E. F. *J. Clim. appl. Met.* **23**, 1012–1031 (1984).

Closure phase in high-resolution optical imaging

J. E. Baldwin*, C. A. Haniff†, C. D. Mackay†
& P. J. Warner*

* Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

† Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

The angular resolution of long-exposure optical images taken with large telescopes at the best sites is limited by atmospheric phase fluctuations to ~ 0.5 – 1 arc s; that is, ~ 10 – 50 times worse than the theoretical diffraction limit. In contrast, the measurement of visibilities and closure phases of fringe patterns in short-exposure images, taken through an aperture mask comprising a non-redundant array of three or more holes, offers the prospect of reliable diffraction-limited imaging using methods well established in very long baseline interferometry (VLBI) at radio wavelengths. Here we report successful first measurements of fringe visibility and closure phase by this technique, together with evidence that the method is applicable to objects as faint as magnitude 15.

Various interferometric techniques have been proposed^{1–4} for obtaining high-resolution information from measurements of the spatial coherence function. The amplitude of the function is easily obtained, but the phase is seriously disturbed by atmospheric and instrumental fluctuations. These atmospheric disturbances can be characterized by the spatial scale, r_0 (typically 0.1 m), across which the r.m.s. phase fluctuation is 1 radian, and the coherence time, τ_c (typically 10 ms), over which the r.m.s. phase fluctuation at a fixed point is 1 radian. In the absence of useful phase measurements, the reconstruction of reliable images is difficult and sometimes impossible. Algorithms for extracting phase information from short-exposure images⁵ break down in conditions of poor signal-to-noise ratio. Except for simple cases such as binary stars, or where a bright star at one edge of the image allows speckle holography to be used⁶, no reliable diffraction-limited optical images have been obtained by any of these techniques. In radioastronomy, on the other hand, diffraction-limited imaging is achieved routinely in VLBI under conditions in which phase disturbances produced by the atmosphere are the same as those experienced in optical imaging, differing only in their spatial and temporal scales.

The simplest form of VLBI array comprises three telescopes. Correlators are used on each of the three interferometer baselines to provide measures of the amplitude and phase of the spatial coherence function. The true phases ϕ_{12} , ϕ_{23} and ϕ_{31} on the three baselines are disturbed by the unknown atmospheric and instrumental phases α_1 , α_2 and α_3 at the three telescopes to give the measured values $(\phi_{12} + \alpha_1 - \alpha_2)$, $(\phi_{23} + \alpha_2 - \alpha_3)$ and $(\phi_{31} + \alpha_3 - \alpha_1)$. The sum of the three measured phases, the closure phase, is $(\phi_{12} + \phi_{23} + \phi_{31})$ and is a property of the coherence function alone, independent of the atmospheric and instrumental effects. The methods for using this type of information for constructing images are reviewed in ref. 7.

The optical analogue of this interferometer consists of three holes, each smaller than the scale, r_0 , of the atmospheric fluctuations, in a mask placed at the telescope aperture, or equivalently in a re-imaged pupil plane. The image of an unresolved object is in effect the diffraction pattern of a single hole crossed by three sets of interference fringes corresponding to the separations and orientations of the three holes. The phases of the fringe patterns can be determined relative to an arbitrary point in the image plane; the closure phase is independent of the location of this point. For a point source the closure phase is always zero, and this provides a convenient test to demonstrate the technique.

The application of closure phase measurement to optical interferometry was first suggested by Rogstad⁸ and the method

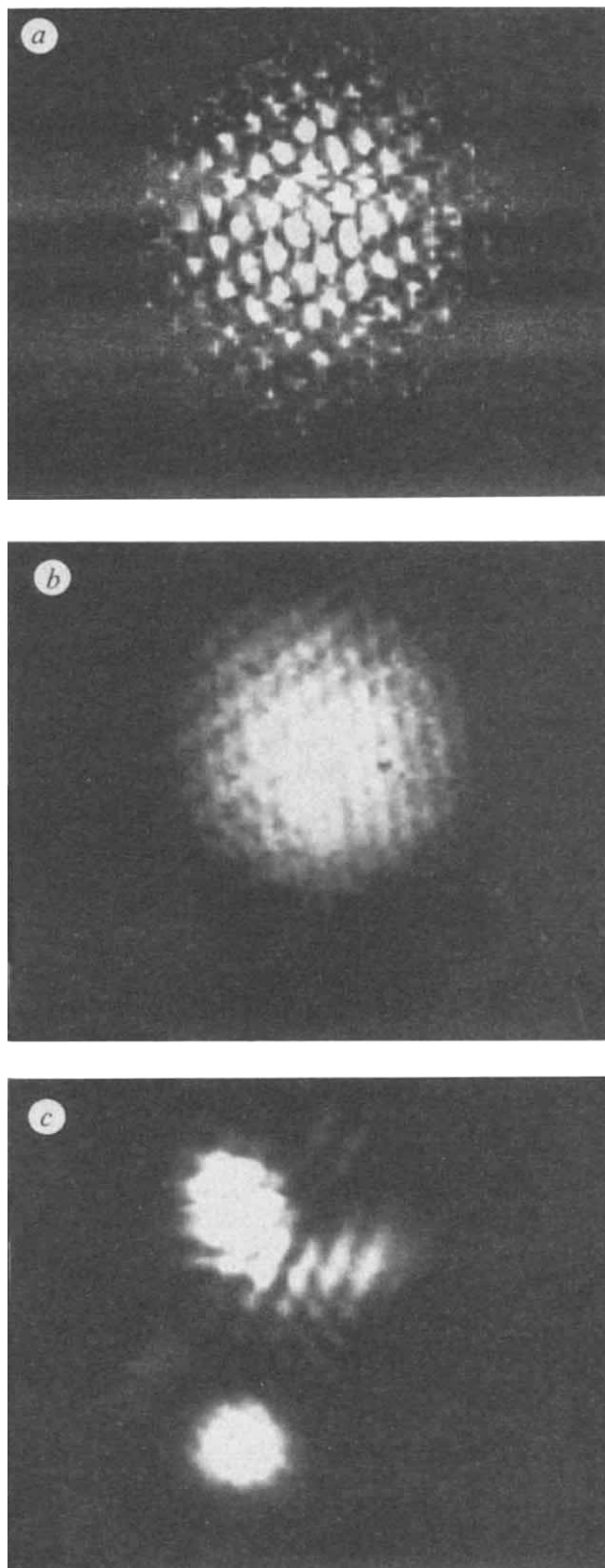


Fig. 1 Optical images obtained with three-hole aperture masks. *a*, β Ori, $m = 0.1$; 13.5 cm effective aperture diameter; 50 ms exposure. *b*, As *a*; 1 s exposure. *c*, β Tau, $m = 1.6$; 45 cm effective diameter; 100 ms exposure; slightly defocused, which separates the images from the three apertures without significantly defocusing the individual images.

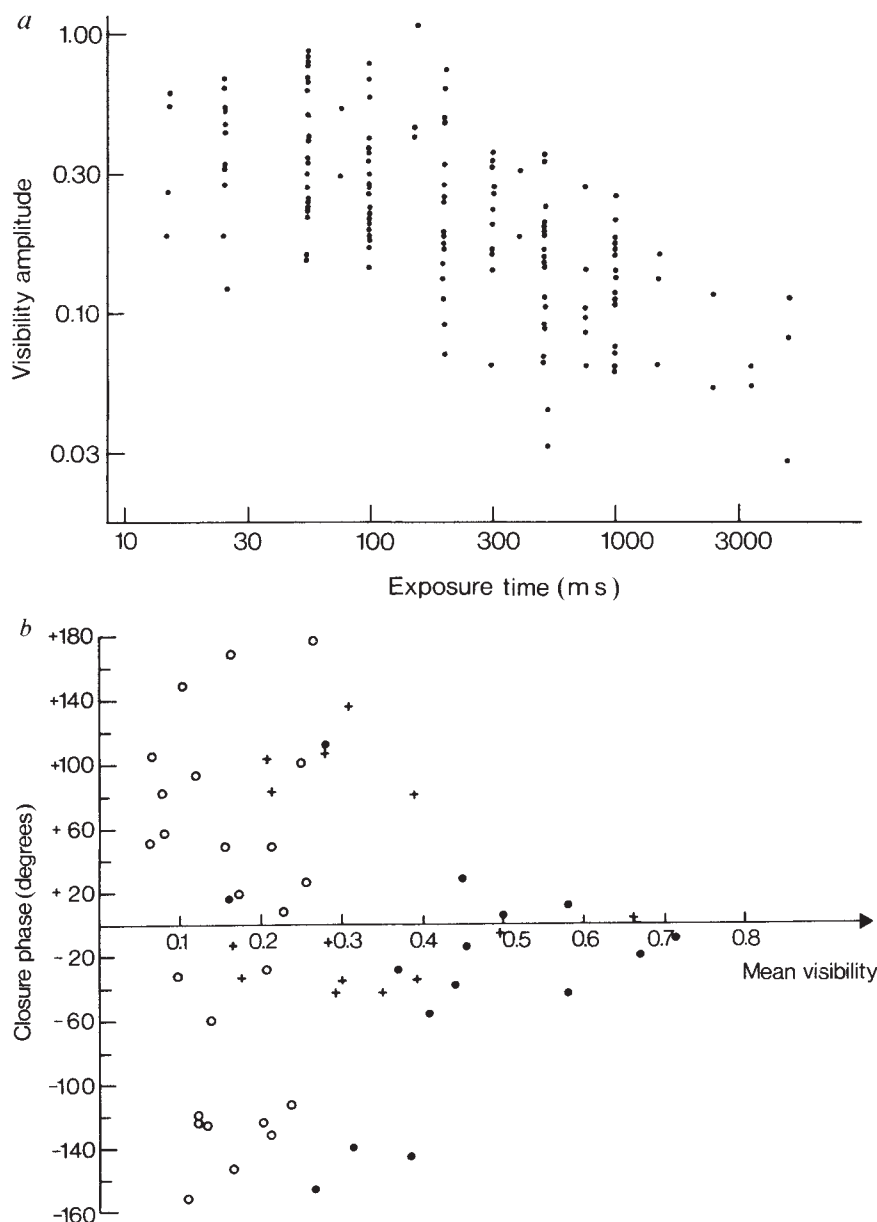


Fig. 2 *a*, Observed fringe visibilities of unresolved stars as a function of exposure time. *b*, Dependence of the measured closure phases on the mean visibility of the fringes in an exposure. Symbols represent different ranges of exposure times: ●, <50 ms; +, 50–200 ms; ○, >200 ms.

described here using non-redundant aperture masks was discussed by Rhodes and Goodman⁹. A mistaken belief that the sensitivity of the method is very poor may explain why it has not been used in practice.

Images were obtained on the night of 4 December 1984 with the Galileo/Institute of Astronomy CCD (charge-coupled device) camera on the University of Hawaii 88-inch telescope on Mauna Kea. The telescope pupil was re-imaged behind the $f/10$ Cassegrain focus with a 50-mm $f1.4$ lens; at the pupil plane the diameter of the pupil was 5.0 mm. Another lens focused the parallel beam onto the CCD so that the final image scale was 36.5 pixels per arc s, giving adequate sampling for baseline lengths up to 1.7 m. An interference filter centred on 6,770 Å (FWHM 155 Å) defined the bandwidth. Several different non-redundant aperture masks were placed in the pupil plane. We used between 3 and 6 apertures per mask, with effective aperture sizes of 13.5–45 cm and effective spatial separations of the apertures of 0.45–2 m. Exposure times varied from 17 ms to 5 s. Observing conditions were average for this site, with some thin cloud, wind speeds of 15 km h⁻¹ and wind shear of ~45° between the upper and lower layers of the atmosphere. The seeing, in the conventional sense, is estimated to have been between 1 and 1.5 arc s (FWHM).

Figure 1*a, b* shows typical CCD frames of a single, bright, unresolved star (magnitude $m = 0.1$), taken using a 3-hole mask with effective hole diameters of 13.5 cm. Three sets of fringe patterns, each corresponding to a particular pair of apertures, are seen superimposed on the overlapping seeing disks of the source.

The amplitude and phase of each fringe pattern were obtained by Fourier transformation of the image frames. For baseline lengths of 0.45–2 m we find no apparent correlation between visibility amplitude and baseline, and we have plotted the visibility amplitudes for all spacings against exposure time in Fig. 2*a*. The scatter and decrease in amplitude A with exposure time t_{exp} is consistent with a Rayleigh distribution of random-phaser sums ($A \propto t_{\text{exp}}^{1/2}$) characterized by a coherence time $t_c < 50$ ms.

Closure phases were derived from the sum of the visibility phases for all exposure times <200 ms and for a few of the longer exposures. As expected, there was a very large scatter in the closure phases for frames with $t_{\text{exp}} \gg t_c$, which was the case for many of our exposures. However, for $t_{\text{exp}} \leq 100$ ms, 64% of the frames gave closure phases <40°, and for the frames with the highest visibilities (such as Fig. 1*a*) the closure phases were <10°, consistent with the zero closure phase expected for an unresolved source. The scatter in closure phase measurements

as a function of the observed fringe visibility is shown in Fig. 2b.

Figure 1c shows an exposure where the effective aperture size has been increased to 45 cm. This frame is slightly defocused so that the images of the star through each of the apertures are slightly displaced from one another. The main peaks have widths (FWHM) of 0.3–0.4 arc s, corresponding closely to the diffraction limit of the apertures (0.33 arc s). Some light is scattered over larger angles. Differing wavefront tilts across the individual apertures give rise to relative movements of the centroids of the three images from frame to frame of 0.2 arc s r.m.s. For $t_{\text{exp}} > 100$ ms the peaks are blurred due to these motions. These results provide a direct measurement of the wavefront curvature on scales of 1–2 m, and hence extend the results obtained with the image-stabilizing instrument system (ISIS) developed at the University of Hawaii¹⁰. Real-time corrections based on such measurements would give diffraction-limited images for apertures > 0.5 m. The data are in reasonable agreement with conventional models of atmospheric phase fluctuations, which predict that most of the power in the spectrum is restricted to the lowest-wavenumber modes. Accurate measurement of the spectrum of the atmospheric fluctuations at different sites is crucial for the optimum design of future imaging experiments.

Practical application of this technique to produce images requires adequate sampling of the spatial coherence function, greater sensitivity in each exposure, and averaging the results of a large number of exposures. Increasing the number of holes in the mask gives a larger number of baselines, which reduces the total observing time for a particular object. It also gives many simultaneous closure phases, which provide a larger fraction of the complete phase information. However, the limiting sensitivity is worse than for a three-hole mask and it is important to make a non-redundant layout of the holes. Redundant baselines both reduce the number of independent phase measurements and combine to give a fringe which in general has low visibility due to the random phases of the two superposed fringes.

The sensitivity in a single exposure would be increased by the use of a photon-counting detector with a very low dark count. The limiting sensitivity for the procedure is determined by the number of photons in a single exposure necessary to define a closure phase with an r.m.s. accuracy better than $\sim 90^\circ$. The average value deduced from N exposures then has an uncertainty which improves as $1/\sqrt{N}$. For an unresolved source and an exposure recording n photons with an m -hole mask, the signal-to-noise ratio of the fringe for each baseline, (amplitude/noise) $\approx (2n/m)/\sqrt{n}$ is ≥ 1 if the mean number of photons per hole in one exposure is $\geq m/4$. Thus, for a small number of holes in the mask, the limiting sensitivity corresponds to a flux of 1–2 photons per hole in each exposure.

Using 0.1-m apertures, 10 ms exposure times, 25 nm bandwidth and an overall photon detection efficiency of the atmosphere, optics and detector of 10%, the limiting magnitude for unresolved objects is magnitude 10.5. The observations described here suggest that, at least on some nights, apertures as large as 0.5 m may be usable with integration times as long as 50 ms, provided that the images from the separate apertures can be superposed. In this case, limiting magnitudes as faint as 15 are attainable.

Properties of discrete electrostatic systems

Hector A. Munera

Swiss Federal Institute of Technology,
Institute of Energy Technology, Nuclear Engineering Laboratory,
ETH-Zentrum, CH-8092 Zurich, Switzerland

Further to a recent controversy^{1–5}, it is well known that continuous classical electrostatics^{6,7} is an approximate description which smoothes the ‘microscopic irregularities’⁷ of discrete systems. Our purpose here is to highlight these microscopic details. We have studied the stability, Coulomb potential energy (W), electrostatic potential (V) and field intensity (E) for several arrangements of N point charges q on and inside the surface of a sphere of radius R ($N \leq 20$). As expected, discrete and continuous configurations differ in several respects. The energy form-factor $w = W/N^2$ for a surface-charged sphere ranges from $0.125 q^2/R$ at $N=2$ to $0.377 q^2/R$ at $N=20$ (the latter value is $\sim 75\%$ of the continuous limit, $0.5 q^2/R$). The field intensity inside the sphere is non-zero in the discrete case, but quickly tends, with increasing N , towards the continuous limit (zero). The common practice of treating small- N physical systems as continuous conceals properties that would otherwise be obvious.

Consider a sphere with centre at the origin of coordinates and let all lengths be in reduced units of R . Equal point charges are located at fixed points P_i , inside and/or on the sphere. Our objective is to find three-dimensional configurations minimizing W (units: q^2/R), which is given by the exact equation⁶

$$W = \sum_{i < j}^N 1/\rho_{ij} = \sum_{i < j}^N 1/|\rho_i - \rho_j| \quad (1)$$

where $\rho_i = |\rho_i| \leq 1$ for all $i = 1, 2, \dots, N$.

The electrostatic potential V (units: q/R) generated by the N -particle system at any point P_0 ($\rho_0, \vartheta_0, \varphi_0$) is

$$V = \sum_{i=1}^N 1/\rho_{i0} = \sum_{i=1}^N 1/|\rho_0 - \rho_i| \quad (2)$$

and the corresponding electrostatic field intensity E (units: q/R^2) is

$$E = \sum_{i=1}^N \rho_{i0}/\rho_{i0}^3 \quad (3)$$

where the spherical coordinates ρ (polar distance), ϑ (angle on equatorial plane $X-Y$) and φ (angle with Z -axis) are related by

$$\rho_{i0}^2 = \rho_i^2 + \rho_0^2 - 2\rho_i\rho_0[\cos \varphi_i \cos \varphi_0 + \sin \varphi_i \sin \varphi_0 \cos (\vartheta_i - \vartheta_0)]$$

A configuration is in rotational equilibrium if the angular field intensity components at each charge, E_ϑ and E_φ , are zero (E is due to the remaining $N-1$ charges.) Examples of such configurations are regular polygons and polyhedra⁸ inscribed in a sphere; in both cases, however, there is a net radial field intensity component ($E_\rho > 0$), to be balanced by forces other than electrostatic.

Rotational equilibrium does not imply minimum energy; for instance, the cube and the dodecahedron are non-optimum configurations (see Table 2). Two-dimensional configurations embedded in three-dimensional space require even greater care. For example, a regular polygon with N vertices (an N -gon) is the minimum- W configuration when the point charges lie on the circumference of a circle⁹; however, if charges are allowed inside the circle, then equilibrium configurations with lower W are possible, as discussed next.

For $N \leq 20$ we have studied all configurations belonging to the family of one- and two-nested regular polygons plus a central charge, fulfilling the following equilibrium conditions: (1) An outer N_0 -gon is inscribed in a great circle of the sphere ($r_0 = 1$),

Received 23 December 1985; accepted 12 February 1986.

1. Michelson, A. A. *Astrophys. J.* **51**, 257–262 (1920).
2. Labeyrie, A. *Astr. Astrophys.* **6**, 85–87 (1970).
3. Hanbury-Brown, R., Davis, J. & Allen, L. R. *Mon. Not. R. astr. Soc.* **137**, 375–392 (1967).
4. Roddier, C. & Roddier, F. in *IAU Colloq.* **49**, 175–185 (1978).
5. Knox, K. T. & Thompson, B. J. *Astrophys. J.* **193**, L45–L48 (1974).
6. Weigelt, G. & Baier, G. *Astr. Astrophys.* **150**, L18–L20 (1985).
7. Pearson, T. J. & Readhead, A. C. S. *A. Rev. Astr. Astrophys.* **22**, 97–130 (1984).
8. Rogstad, D. H. *Appl. Opt.* **7**, 585–588 (1968).
9. Rhodes, W. T. & Goodman, J. W. *J. opt. Soc. Am.* **63**, 647–657 (1973).
10. Thompson, L. *Astrophys. J.* **279**, L47–L49 (1984).

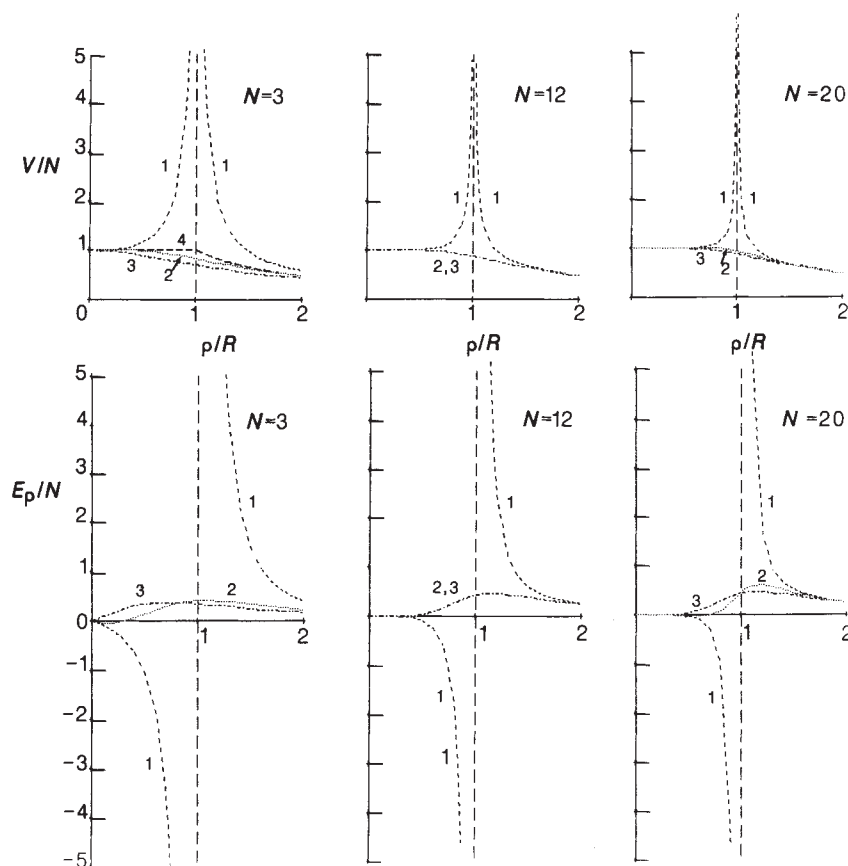


Fig. 1 Coulomb potential and radial field intensity in three-dimensional discrete electrostatic configurations. A point charge q is located at each vertex of an equilateral triangle ($N=3$), an icosahedron ($N=12$) and a dodecahedron ($N=20$), inscribed in a sphere (radius R). Potential (above) and radial field (below) per unit charge along three symmetry axes are plotted as curves 1–3. Potential along vertex-rays (through the point charges; curve 1) becomes infinite at the surface of the sphere; the corresponding radial field intensity switches from repulsive (outside) to attractive (inside the sphere). The radial field intensity along edge-rays (through midpoints of edges; curve 2) is attractive (negative) close to the centre; the field along face-rays (through centres of faces; curve 3) is always positive, vanishing at the centre. The potential per unit charge produced by a uniformly charged hollow sphere is q/R inside and q/ρ elsewhere (curve 4); the corresponding inner field intensity is zero. The discrete curves approach the classical continuous treatment at $\sim 2R$.

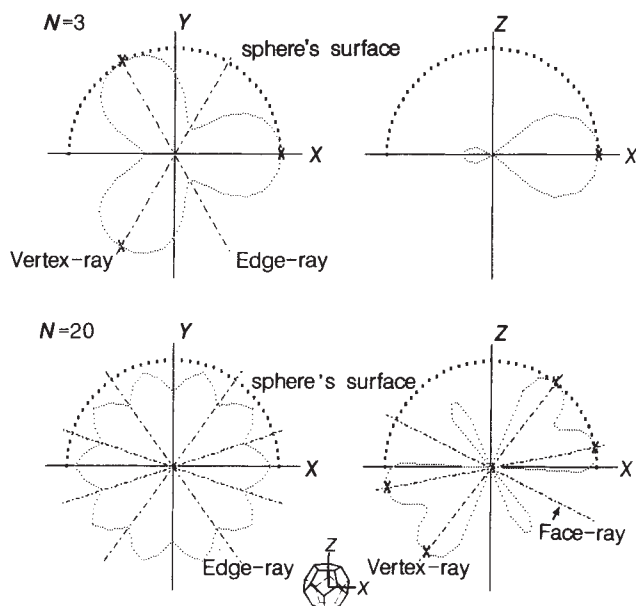


Fig. 2 Surfaces of three-dimensional discrete electrostatic configurations. Along each direction, V reaches a maximum at ρ^* such that $E_\rho = 0$ (except for vertex-rays; see Fig. 1). The locus ρ^* (faint dotted line) is the surface of the configuration. Above, two cross-sections of the electrostatic surface of an equilateral triangle contained in the X - Y plane. Below, two cross-sections for the dodecahedron; its orientation is shown in the inset, with the Y -axis perpendicular to the page. Along some directions the surface is zero, leading to 'classical transparency'. Conventionally, the surface of the configuration is the sphere (bold dotted line).

in rotational equilibrium. (2) A coplanar inner N_i -gon is inscribed in a concentric circle ($r_i < 1$, $N_i \leq N_0$), in stable rotational and radial equilibrium. Table 1 shows the optimum configuration for each N . As noted by Berezin¹, a charge at the centre is optimum for $N > 11$. For $N > 17$, however, minimum W obtains at progressively more complex configurations. For instance, for $N=100$ there are 12 stable equilibrium configurations; minimum energy obtains for the partition 66/33/1 ($W = 6,557.7$, much lower than for Berezin's 99/1 configuration ($W = 7,462.8$) or the 100-gon ($W = 7,529.3$)).

Although the configurations of Table 1 are stable in two dimensions, they do not exhibit stability in three. Hence, they are not absolute optima, as can be seen by comparing Table 1 with our best three-dimensional configurations in Table 2. Furthermore, even in two dimensions, Queen² reports for $N=17$ a non-equilibrium configuration with energy lower than the corresponding minimum in Table 1.

We turn now to the strictly three-dimensional problem of N charges on the surface of a sphere. The equatorial plane is in the X - Y plane, and the north and south poles are on the Z -axis. The charges are partitioned among the vertices of regular polygons, parallel to the equatorial plane. Seven groups have been analysed so far: (1) An equatorial polygon (EP) with n_e charges plus one charge at each (north and south) pole. Notation: bicapped EP: $1/n_e/1$, where $N = n_e + 2$. (2) Two n -gons ($n = N/2$) symmetrically disposed at distances h_2 on either side of the equatorial plane. In reality this is a prism denoted as P: n/n . An antiprism (AP: n/n) is obtained by rotating the southern n -gon through an angle π/n . (3) Similar to configuration (2) plus one charge at each pole. Notation: bicapped (A)P: $1, n/n, 1$ where $N = 2(n+1)$. (4) One equatorial n_e -gon plus an (anti)prism. Notation: (A)P, EP: $n/n_e/n$, where $N = n_e + 2n$. (5) Similar to configuration (4) plus one charge at each pole. Notation: bicapped (A)P, EP: $1, n/n_e/n, 1$, where $N = n_e + 2n + 2$. (6) Two nested coaxial (anti)prisms whose bases are n_1 -gons

Table 1 Minimum energy planar configurations under stable equilibrium in the class of one- and two-nested polygons plus central charge

Total N	Charges		Central	Inner radius* r_i	Rotation angle β	Form-factor $w(N)$	Potential energy $W(q^2/R)$
	Outer N_o	Inner N_i					
2	2	—	—	—	—	0.125	0.50
3	3	—	—	—	—	0.192	1.73
4	4	—	—	—	—	0.239	3.83
5	5	—	—	—	—	0.275	6.88
6	6	—	—	—	—	0.305	10.96
7	7	—	—	—	—	0.329	16.13
8	8	—	—	—	—	0.351	22.44
9	9	—	—	—	—	0.369	29.92
10	10	—	—	—	—	0.386	38.62
11	11	—	—	—	—	0.401	48.58
12	11	—	1	—	—	0.414	59.58
13	12	—	1	—	—	0.425	71.81
14	13	—	1	—	—	0.435	85.35
15	14	—	1	—	—	0.445	100.22
16	15	—	1	—	—	0.455	116.45
17	16	—	1	—	—	0.464	134.06
18	16	2	—	0.304	$\pi/16$	0.471	152.49
19	16	2	1	0.49	$\pi/16$	0.480	173.40
20	18	2	—	0.293	$\pi/18$	0.485	194.02

* In the numerical computation, step sizes of 0.01 and 0.001 were used. This affects W at most in the last digit.

Table 2 Minimum Coulomb energy three-dimensional configurations for $4 \leq N \leq 20$

Total charges N	Configuration*	Distances from origin		Rotation angle β	Form-factor $w(N)$	Potential energy $W(q^2/R)$
		h_1	$h_2^\dagger$			
4	AP: 2/2	NA	$\sqrt{3}/3$	0	0.22964	3.674§
5	Bicapped EP: 1/3/1	NA	NA	0	0.25899	6.475
6	AP: 3/3	NA	$\sqrt{3}/3$	0	0.27737	9.985§
7	Bicapped EP: 1/5/1	NA	NA	0	0.29496	14.453
8	AP: 4/4	NA	0.56	0	0.30743	19.675
	Bicapped AP: 1,3/3,1	NA	1/3	0	0.30845	19.741§
9	P, EP: 3/3/3	0.70	NA	$\pi/3$	0.31802	25.760
10	Bicapped AP: 1,4/4,1	NA	0.42	0	0.32717	32.717
11	Bicapped P, EP: 1,3/3/3,1	0.53	NA	$\pi/3$	0.33572	40.622
12	AP, AP: 3,3/3,3	0.79	0.19‡	$\pi/3$	0.34143	49.165§
13	AP, EP: 4/5/4	0.75	NA	$\pi/40$	0.34946	59.058
14	Bicapped AP: 1,6/6,1	NA	0.46	0	0.35360	69.306
15	AP, EP: 4/7/4	0.78	NA	$\pi/56$	0.35893	80.759
16	AP, AP: 4,4/4,4	0.78	0.20	$\pi/4$	0.36297	92.920
17	Bicapped P, EP: 1,5/5/5,1	0.61	NA	$\pi/5$	0.36696	106.050
18	Bicapped AP, AP: 1, 4,4/4,4,1	0.68	0.20	$\pi/4$	0.37063	120.084
19	Bicapped AP, EP: 1,5/7/5,1	0.67	NA	$\pi/70$	0.37441	135.161
20	Bicapped P, AP: 1,6,3/3,6,1	0.62	0.00	$\pi/6$	0.37727	150.908
	AP, AP: 5,5/5,5	0.79	0.19‡	0	0.37950	151.799§

* AP, antiprism; P, prism; DP, dipyrmaid; EP, equatorial polygon.

† Typical numerical accuracy for optimum h : 0.0001. Accuracy in W at least as good. For odd- N : odd n_e -gon on the equatorial plane, hence h_2 is not applicable (NA).

‡ Exactly: $h_1 = [75 + 30\sqrt{5}]^{1/2}/15$ and $h_2 = [75 - 30\sqrt{5}]^{1/2}/15$.

§ Platonic solid. Note that cube and dodecahedron are not optimum.

and n_2 -gons located at distances h_1 and h_2 from the equatorial plane, such that $h_1 > h_2$. Notation: (A)P, (A)P: $n_1, n_2/n_2, n_1$, where $N = 2(n_1 + n_2)$. (7) Similar to configuration (6) plus one charge at each pole. Notation: bicapped (A)P, (A)P: 1, $n_1, n_2/n_2, n_1, 1$, where $N = 2(n_1 + n_2 + 1)$. The relative rotation angle between the two (anti)prisms or between the (anti)prism and the equatorial polygon is denoted by β . For our purposes, we prefer the present notation to other alternatives⁹⁻¹². Table 2 shows, for each N , the configuration with minimum energy (obtained by a numerical optimization procedure, in which h_1, h_2 and β were varied for all partitions of N among the configuration groups).

As in other extremal problems which involve vertices of polyhedra^{13,14}, our optimal configurations have triangular faces

(Table 2). In particular, the cube (square faces) and the dodecahedron (pentagonal faces) are non-optimal. All configurations in Table 2 are in rotational equilibrium except $N = 13, 15$ and 19.

Our simplified approach reproduces the best configurations reported in the literature for $N \leq 16$ (refs 11, 12, 14-20), excluding $N = 11$, for which a slightly better configuration is known¹⁷ ($W = 40.596$). For $N = 13$ and 15, different configurations have been reported¹², but not the corresponding W s. Our configurations for $N = 17-20$ are believed to be new. For $N = 8$ our isosceles antiprism improves on Fejes¹⁴ equilateral antiprism ($W = 19.725$). For $N = 20$ we improve on both the dodecahedron and Foppl's¹⁵ configuration ($W = 152.163$).

We now list some properties of discrete electrostatic configur-

ations and note some possible physical applications. The Coulomb field intensity inside the sphere is typically non-zero: the angular components (E_θ and E_ϕ) vanish along the symmetry axes, but the radial component (E_ρ) does not (see Fig. 1). Note that $E_\rho \rightarrow 0$ at the centre; the size of the approximately null region increases with N , eventually leading to the continuous classical limit (zero field everywhere).

According to classical electrostatics⁶, a charge Nq uniformly smeared on the surface of a sphere has potential $V = Nq/R$ on and inside the sphere. In the discrete case, however, this expression holds true only at the centre: elsewhere, V is different (see Fig. 1). Also, V exhibits a marked directional dependence; potentials along some directions are reminiscent of some phenomenological potentials used in nuclear physics: hollow spheres, charge concentrated at the origin, volume-charged spheres²¹ and truncated Coulomb potentials²²⁻²⁴.

The discrete energy form-factor w is significantly lower than the classical limits for surface-charged (0.5) and volume-charged (0.6) spheres²¹ (see Tables 1, 2). This may have implications for semi-empirical nuclear-mass formulae for light nuclei, and for nuclear radii²⁵.

Let the effective surface of the discrete configuration be defined as the locus ρ^* where V attains a maximum along a particular direction; its shape is not spherical, as illustrated in Fig. 2. In head-on collisions between particles of charge Q and stationary discrete configurations, the apparent surface depends on the initial energy of the projectile. This surface steadily shrinks with increasing energy until the hard-core (ρ^*) obtains. A suitable averaging procedure on ρ^* may lead to an equivalent radius for rotating systems (possible connections: nuclear radii^{25,26}, proton hard-core, nuclear scattering). A remarkable fact in Fig. 2 is that $\rho^* = 0$ along some directions. This implies that the configuration may be transparent to a charged particle travelling with non-relativistic energy $K > NqQ/R$ (with q and Q of the same sign). If the configuration is rotating, there is still a finite probability that the particle may appear undisturbed at the other side of the system (possible connections: nuclear scattering). The converse phenomenon is *classical tunnelling*; that is, the possibility that a charge Q originally at rest in the centre of the sphere (unstable equilibrium) may escape along some preferred directions, appearing at infinity with kinetic energy $K = NqQ/R$ (possible connections: radioactive decay).

The two-dimensional symmetry of stable nested polygons (Table 1) is quite remarkable (possible connections: two-dimensional crystal growth processes, as in snowflakes²⁷⁻²⁹). The energetically favourable three-dimensional configurations in Table 2 often exhibit non-pentagon symmetries ($N = 20$ is noteworthy); even the icosahedron, paradigm of 5-fold symmetry, may be described as two-nested triangular antiprisms¹⁴ as shown in Table 2 (possible connections: three-dimensional crystals).

We thank A. A. Berezin, O. Knop, N. M. Queen and our colleagues B. Sigg, K. Tichy and G. Yadigaroglu for comments and suggestions.

15. Foppl, L. *J. reine angew. Math.* **141**, 251-302 (1912).
16. Goldberg, M. *Maths Comput.* **23**, 785-786 (1969).
17. Cohn, H. *Mathl. Tabl. natn. Res. Council, Wash.* **10**, 117-120 (1956).
18. Lin, Y. C. & Williams, D. E. *Can. J. Chem.* **51**, 312-316 (1973).
19. Claxton, T. A. & Benson, G. C. *Can. J. Chem.* **44**, 157-163; 1730-1731 (1966).
20. Britton, D. *Can. J. Chem.* **41**, 1632-1634 (1963).
21. Evans, R. D. *The Atomic Nucleus*, 32-39 (McGraw-Hill, New York, 1955).
22. Singh, D., Varshni, Y. P. & Dutt, R. *Phys. Rev. A* **32**, 619-622 (1985).
23. Patil, S. H. *Phys. Rev. A* **24**, 2913-2919 (1981).
24. Mehta, C. H. & Patil, S. H. *Phys. Rev. A* **17**, 43-46 (1978).
25. Heifer, E. F. *Phys. Rev. A* **32**, 1205-1027 (1985).
26. Elton, L. R. B. *Nuclear Sizes* (Oxford University Press, 1961).
27. Maddox, J. *Nature* **313**, 93 (1985).
28. Mortley, W. S. *Nature* **313**, 638 (1985).
29. Schrack, R. A. *Nature* **314**, 324 (1985).

Lunar nodal tide and distance to the Moon during the Precambrian

James C. G. Walker* & Kevin J. Zahnle*†

* Department of Atmospheric and Oceanic Sciences, and
† Department of Astronomy, The University of Michigan,
Ann Arbor, Michigan 48109, USA

The pace of tidal evolution for the past ~450 Myr implies an Earth/Moon collision some 1,500-2,000 Myr BP (see ref. 1), an event for which there is no corroborating evidence. Here we present the first direct determination of the lunar distance in the Precambrian. We interpret a 23.3 ± 0.3 -yr periodicity preserved in a 2,500 Myr BP Australian banded iron formation (BIF)² as reflecting the climatic influence of the lunar nodal tide, which has been detected with its modern 18.6-yr periodicity in some modern climate records³⁻¹⁰. The lunar distance at 2,500 Myr BP would then have been about 52 Earth radii. The implied history of Precambrian tidal friction is in accord with both the more recent palaeontological evidence and the long-term stability of the lunar orbit. The length of the Milankovitch cycles that modulate the ice ages today¹¹⁻¹³ also evolve with the Earth-Moon system. Their detection in the Precambrian sedimentary record would then permit an independent determination of the lunar distance.

Evidence for the rate of lunar recession is available on several different timescales. As these matters have been discussed in detail elsewhere^{1,14,15}, we shall discuss them only briefly here. Direct determination by laser ranging yields a recession velocity $\dot{a} \approx 1.1 \times 10^{-7}$ cm s⁻¹ (ref. 16), where a refers to the lunar distance. The amplitudes and phases of the various tides can be inferred from their influences on satellite orbits, from which $\dot{a} \approx 1.2 \times 10^{-7}$ cm s⁻¹ (ref. 16) is calculated. Over longer periods, lunar occultation timings and the historical record of total solar eclipses imply that $\dot{a} \approx 1.3 \times 10^{-7}$ cm s⁻¹ over the past few millennia¹.

The most important source of long-term information derives from fossil corals and molluscs^{1,17}. In brief, fine laminae are interpreted either as daily growth increments or as the record of the semidiurnal or diurnal tide, in accordance with the general growth habits of their modern descendants. Modulation of the fine banding by the fortnightly or monthly tidal cycles and by the yearly seasonal cycle allows estimates of the number of days per year, the number of days per month, and the number of months per year. Lambeck's preferred solution, obtained by considering a filtered set of the palaeontological data and additionally imposing conservation of angular momentum for the Earth-Moon system, may be expressed as $\dot{a} \approx 9.5 \times 10^{-8}$ cm s⁻¹. Scrutton's evaluation of the palaeontological data is essentially identical¹⁷.

The quality of the palaeontological record declines abruptly when looking back into the Precambrian. The only previously exploited data come from stromatolites. However, stromatolites

Received 29 October 1985; accepted 11 February 1986.

1. Berezin, A. A. *Nature* **315**, 104 (1985).
2. MacGowan, D. *Nature* **315**, 635 (1985).
3. Nityananda, R., Cormack, A. M., Naumann, R. A. & Webb, S. *Nature* **316**, 301-302 (1985).
4. Queen, N. M., Rees, M. & Berezin, A. A. *Nature* **317**, 208 (1985).
5. Aspdin, H. *Nature* **319**, 8 (1986).
6. Jeans, J. H. *The Mathematical Theory of Electricity and Magnetism* 5th edn, 36, 42 (Cambridge University Press, 1933).
7. Morse, P. M. & Feshbach, H. *Methods of Theoretical Physics* Vol. 1, 201 (McGraw-Hill, New York, 1953).
8. Leech, J. *Mathl. Gaz.* **41**, 81-90 (1957).
9. Fejes Toth, L. *Regular Figures*, 102-123, 157 (Pergamon, Oxford, 1964).
10. Coxeter, H. S. M. *Regular Polytopes*, 33-57 (Methuen, London, 1948).
11. King, R. B. *J. Am. chem. Soc.* **92**, 6455-6466 (1970).
12. Melnyk, T. W., Knop, O. & Smith, W. R. *Can. J. Chem.* **55**, 1745-1761 (1977).
13. Fejes Toth, L. *Am. J. Math.* **70**, 174-180 (1948).
14. Whyte, L. L. *Am. math. Mont.* **59**, 606-611 (1952).

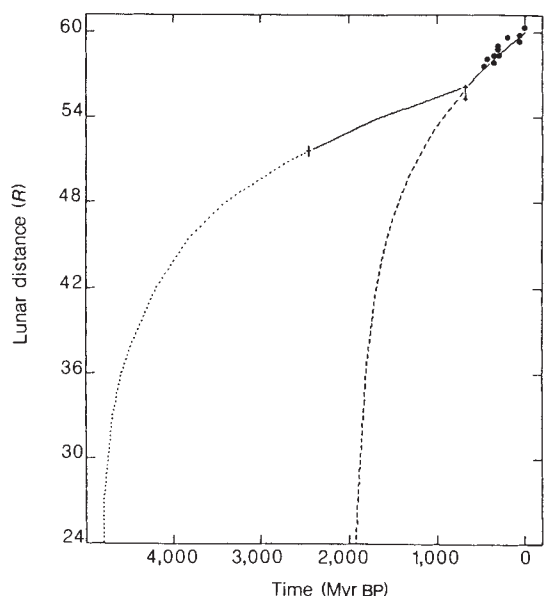


Fig. 1 Radius of the lunar orbit (in units of R , the Earth radius) through Earth's history. The Phanerozoic data are adapted from ref. 17. The Precambrian datum has been deduced from a 23.3-yr periodicity preserved in the Weeli Wolli banded iron formation, which we have identified with the lunar nodal tide. Dashed curve, an extrapolation [equation (3)] based solely on the Phanerozoic data; dotted curve [equation (5)] presumes an average Precambrian tidal torque consistent with both the palaeontological data (●) and the Weeli Wolli datum (+).

do not appear to be good timekeepers today, and the correct interpretation of ancient stromatolitic data is far from obvious^{1,17}.

Precambrian inorganic sources of periodic banding do exist; however, since inorganic processes are comparatively slow, observable periodicities must be relatively long. The most interesting of these was found by Trendall² in the Weeli Wolli Formation of the Hamersley Group in Western Australia¹⁸. Here it suffices to note that the Weeli Wolli is a banded iron formation dating from ~2,450 Myr BP. The finest bands (or 'varves') are believed to be annual^{2,18}. The varves themselves vary cyclically with a period of 23.3 ± 0.3 yr, counted over 23 complete cycles. The absence of a readily discernible 11-yr cycle led Trendall to doubt that the 23.3-yr cycle is in fact the 22-yr Hale (double-sunspot) cycle.

We tentatively interpret the 23.3-yr cycle as a consequence of the climatic impact of the lunar nodal tide. The lunar nodal tide, presently 18.6 yr, appears weakly in some modern climate records. It arises from the precession of the moon's orbital plane about the ecliptic. It has been detected in temperature and rainfall records in western North America³⁻⁵ and eastern Australia⁵, and in tree-ring data from both western North America^{4,6} and Patagonia⁷. It has also been detected in the Indian monsoons^{8,9} and perhaps in Beijing's rainfall¹⁰ (a claim disputed by Clegg and Wigley¹⁹) and Nile River flooding²⁰.

What is most important from our present perspective is that this period depends on the Earth/Moon distance. The lunar nodal period may be conveniently expressed as²¹

$$P = P_0 \left[\frac{\cos I_0}{\cos I} \right] \left(\frac{a_0}{a} \right)^{1.5} \quad (1)$$

where the modern values are $P_0 = 18.6$ yr, $a_0 = 3.84 \times 10^{10}$ cm, and $I_0 = 5.15^\circ$, which is the inclination of the lunar orbit to the ecliptic.

Interpreted as the lunar nodal period, the 23.3-yr cycle places the Moon at $a/a_0 = 0.86 \pm 0.01$. When plotted in comparison

with the modern rate of tidal dissipation (Fig. 1), it is clear that this is about where the Moon should have been at that time if the catastrophic close approach is to be averted.

Neglecting the solar tide and the inclination and eccentricity of the Moon's orbit, the recession of the Moon due to the semidiurnal tide may be written after Lambeck¹ as

$$\dot{a} = Aa^{-11/2} = \langle \dot{a}_0 \rangle \left(\frac{a}{a_0} \right)^{-11/2} \quad (2)$$

where

$$A = \frac{3}{16} (1 + \cos(i))^4 \frac{GmR^5 k_2}{(G(M+m))^{0.5} \sin \delta}$$

and where $\langle \dot{a}_0 \rangle$ is the average velocity of lunar recession at $a = a_0$; i the inclination of lunar orbit to Earth's Equator (23.5° today); R and M the radius and mass of Earth; m the lunar mass; G the gravitational constant; and k_2 and δ the equivalent Love number and phase lag of the tide. Assuming that A is constant (equivalent to assuming that tidal energy dissipation is directly proportional to the total tidal energy; that is, assuming a constant Q factor), equation (2) is readily integrated:

$$\frac{a}{a_0} = \left[1 - \frac{13}{2} T \frac{\langle \dot{a}_0 \rangle}{a_0} \right]^{2/13} \quad (3)$$

where T is time measured back from the present. In terms of T :

$$T = \frac{2}{13} \frac{a_0}{\langle \dot{a}_0 \rangle} \left[1 - \left(\frac{a}{a_0} \right)^{13/2} \right] \quad (4)$$

For a lunar recession velocity $\langle \dot{a}_0 \rangle = 1 \times 10^{-7}$ cm s⁻¹, an Earth/Moon collision is indicated 1,900 Myr ago (Fig. 1). As we have only the single datum for the first ~4,000 Myr of Earth history, deriving a single Precambrian tidal dissipation rate to go with the single Phanerozoic rate is as much analysis as can be justified. An equation matching the Precambrian datum to equation (3) is:

$$\frac{a}{a_0} = \frac{a_1}{a_0} \left[1 - \frac{13}{2} (T - T_1) \frac{\langle \dot{a}_1 \rangle}{a_1} \right]^{2/13} \quad (5)$$

where

$$\langle \dot{a}_1 \rangle = 0.30 \langle \dot{a}_0 \rangle \left[\frac{a_1}{a_0} \right]^{-11/2} = (0.43) \langle \dot{a}_0 \rangle$$

and where T_1 refers to the nominal boundary between the Precambrian and the Phanerozoic. Hence, the average tidal friction (relative to the forcing) in the Proterozoic was ~30% that in the Phanerozoic. Tidal friction in the Archaean, although no larger, may have been lower still. This rate of tidal dissipation is consistent with the stability of the lunar orbit, as shown in Fig. 1.

The Milankovitch cycles of ~23, ~41 and ~105 kyr are periodicities of Earth's orbital parameters that have a prominent role in modulating the ice age climate^{12,13}. The longest period is in the magnitude of the eccentricity^{22,23}, which should be constant over the life of the Solar System. By contrast, both the 41-kyr ('obliquity') and the 23-kyr ('precession') cycles depend in part on the precession of Earth's polar axis with respect to the fixed stars (precession of the equinoxes)^{22,23}, which results from lunar and solar torques on Earth's equatorial bulge. Its period may be written as²⁴

$$P_{\text{eqnx}} = 25,800 \left[\frac{\Omega_0}{\Omega} \right] \left[\frac{\cos \gamma_0}{\cos \gamma} \right] \left[\frac{1.465}{(a_0/a)^3 + 0.465} \right] \text{ yr} \quad (6)$$

where γ is the average obliquity of the polar axis, and Ω the rotation angular velocity. For the present argument, we neglect the small evolutionary change in $\cos \gamma$ (ref. 25), and we assume that the Earth/Moon system angular momentum is constant, so

that P_{eqnx} may be written entirely in terms of a :

$$P_{eqnx} = 37,800 \left[\left\{ 5.87 - 4.87 \left(\frac{a}{a_0} \right)^{0.5} \right\} \left\{ \left(\frac{a_0}{a} \right)^3 + 0.465 \right\} \right]^{-1} \quad (7)$$

Since P_{eqnx} varies not only as a function of lunar distance but also as a function of the daylength through its impact on the equatorial bulge, it is possible, in principle, to determine the daylength directly.

The period of obliquity oscillations may be written as

$$P_{obl} = \left[\frac{1}{P_{eqnx}} - \frac{1}{P_{incl}} \right]^{-1} \quad (8)$$

where $P_{incl} = 68.8$ kyr for the most important term²⁴.

The perihelion precession is composed of four larger terms and a myriad of smaller ones^{22,23}. These may be written in the form

$$P_{peri} = \left[\frac{1}{P_{eqnx}} + \frac{1}{P_{ecc}} \right]^{-1} \quad (9)$$

where the most important eccentricity periods are 308, 176, 72.6 and 75.3 kyr (ref. 23). These correspond to P_{peri} of 23.7, 22.4, 19 and 19.2 kyr, respectively.

The amplitude of the obliquity oscillation also varies²⁴. This may be written approximately as

$$\Delta\gamma = 2.0^\circ \frac{P_{eqnx}}{P_{incl} - P_{eqnx}} \quad (10)$$

As there are in fact half a dozen other periods associated with non-negligible, albeit lesser, inclination variations, equation (10) should really be written as a sum, the terms of which are given by Ward²⁴. What is shown by equation (10) is that the magnitude of the obliquity oscillations is increasing, and that it was about 40% of its present value when the Weeli Wolli Formation was laid down.

We conclude that the Milankovitch periodicities that might be expected in the Weeli Wolli Formation or another formation of comparable age would be ~105, ~17 and ~13 kyr, respectively. The shorter of these might be associated with the alteration between BIF bands and S bands reported by Trendall¹⁸ for the Dales Gorge Member of the Brockman iron formation.

This research was supported in part by NASA grant NAGW-176.

Received 6 November 1985; accepted 20 January 1986.

1. Lambeck, K. *The Earth's Variable Rotation* (Cambridge University Press, New York, 1980).
2. Trendall, A. F. *Econ. Geol.* **68**, 1089–1097 (1973).
3. Currie, R. G. *J. geophys. Res.* **86**, 11055–11064 (1981).
4. Currie, R. G. *J. geophys. Res.* **89**, 1295–1308 (1984).
5. Vines, R. G. *J. geophys. Res.* **87**, 7303–7311 (1982).
6. Currie, R. G. *J. geophys. Res.* **89**, 7215–7230 (1984).
7. Currie, R. G. *Geophys. Res. Lett.* **10**, 1089–1092 (1983).
8. Campbell, W. H., Blechman, J. B. & Bryson, R. A. *J. Clim. appl. Met.* **22**, 289–296 (1983).
9. Currie, R. G. *Geophys. Res. Lett.* **11**, 50–53 (1984).
10. Hameed, S., Yeh, W. M., Cess, R. D. & Wang, W. C. *Geophys. Res. Lett.* **10**, 436–439 (1983).
11. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **1294**, 1121–1132 (1976).
12. Crowley, T. J. *Rev. Geophys.* **21**, 828–877 (1983).
13. Lorius, C. *et al. Nature* **316**, 591–596 (1985).
14. Brosche, P. & Sündermann, J. (eds) *Tidal Friction and the Earth's Rotation* (Springer, New York, 1978).
15. Brosche, P. & Sündermann, J. (eds) *Tidal Friction and the Earth's Rotation* Vol. 2 (Springer, New York, 1982).
16. Cazenave, A. in *Tidal Friction and the Earth's Rotation* Vol. 2 (eds Brosche, P. & Sündermann, J.) 4–18 (Springer, New York, 1982).
17. Scrutton, C. T. in *Tidal Friction and the Earth's Rotation* (eds Brosche, P. & Sündermann, J.) 154–196 (Springer, New York, 1978).
18. Trendall, A. F. in *Iron Formations: Facts and Problems* (eds Trendall, A. F. & Morris, R. C.) 69–129 (Elsevier, New York, 1983).
19. Clegg, S. L. & Wigley, T. M. L. *Geophys. Res. Lett.* **11**, 1219–1222 (1984).
20. Hameed, S. *Geophys. Res. Lett.* **843**–845 (1984).
21. Kaula, W. M. *An Introduction to Planetary Physics* (Wiley, Toronto, 1969).
22. Berger, A. *Astr. Astrophys.* **51**, 127–135 (1976).
23. Berger, A. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
24. Ward, W. R. *Icarus* **50**, 444–448 (1982).
25. Goldreich, P. *Rev. Geophys.* **4**, 411–439 (1966).

Sahel rainfall and worldwide sea temperatures, 1901–85

C. K. Folland, T. N. Palmer & D. E. Parker

Meteorological Office, London Road, Bracknell RG12 2SZ, UK

Using the comprehensively quality-controlled Meteorological Office Historical Sea Surface Temperature data set (MOHSST)^{1,2} we show for the first time that persistently wet and dry periods in the Sahel region of Africa are strongly related to contrasting patterns of sea-surface temperature (SST) anomalies on a near-global scale. The anomalies include relative changes in SST between the hemispheres, on timescales of years to tens of years, which are most pronounced in the Atlantic. Experiments with an 11-level global atmospheric general circulation model (AGCM) support the idea that the worldwide SST anomalies modulate summer Sahel rainfall through changes in tropical atmospheric circulation^{3–6}. El Niño events may also play a part. We do not discount the effects of soil moisture and albedo changes in the Sahel^{7,8}, although Courel *et al.*⁹ have questioned the importance of albedo changes, but we do suggest that worldwide SST anomalies may have a more fundamental influence on Sahel rainfall.

Rainfall records for sub-Saharan North Africa have been collated and normalized by many authors. We have used an

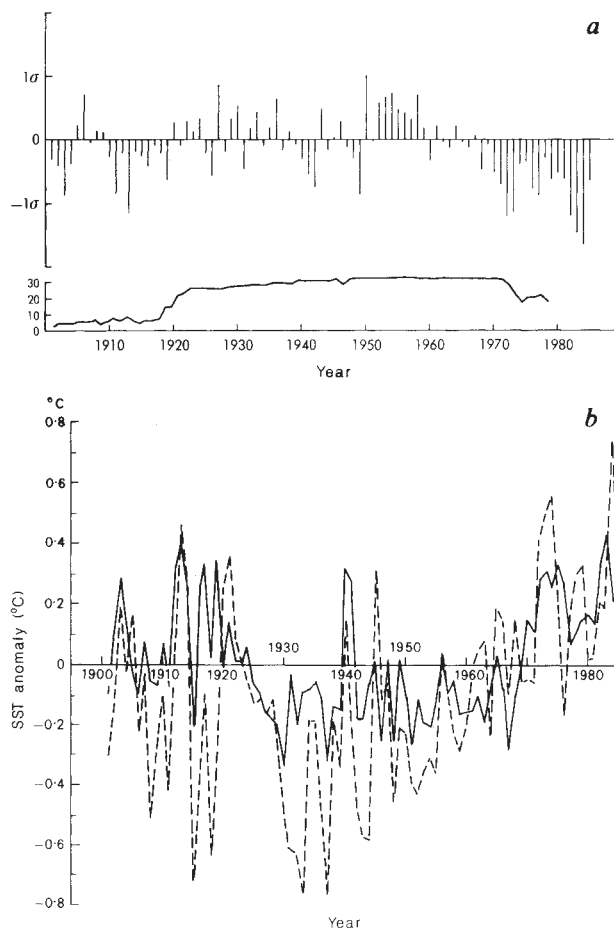


Fig. 1 *a*, Standardized annual rainfall anomalies for the Sahel, 1901–85 (upper panel). Values to 1984 are after Nicholson¹¹; 1985 values are from CLIMAT reports. The lower panel gives the numbers of stations used. *b*, SST anomalies (relative to 1951–80), for July to September 1901–85. Solid line, Southern Hemisphere with the Northern Indian Ocean, minus the rest of the Northern Hemisphere (SHNI – RNH). Dashed line, South Atlantic minus North Atlantic.

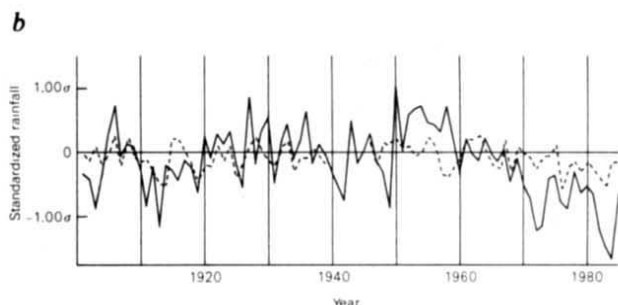
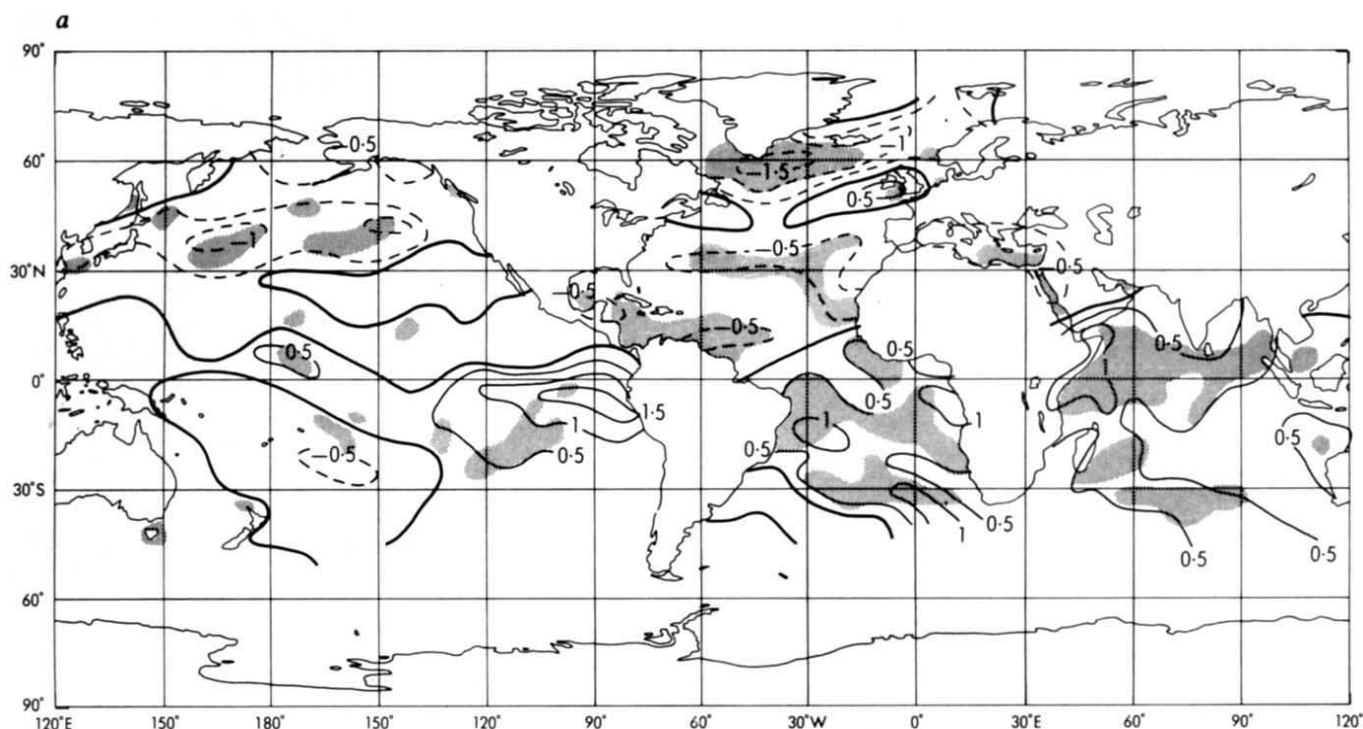


Fig. 2 *a*, SST, July to September: average of (1972–73, 1982–84) (Sahel dry) minus average of (1950, 1952–54, 1958) (Sahel wet). Contours every 0.5 °C. Shaded areas are different from zero at the 90% level of significance according to a *t*-test. *b*, Observed (—) and simulated (---) Sahel rainfall (see text).

updated version of Nicholson's^{10,11} annual Sahel series for 1901–84 (Fig. 1*a*) supplemented by CLIMAT reports for 1985 (see ref. 10 for a map of the area, which stretches from Senegal to Sudan at approximately 14–18° N). Nicholson homogenized her data to allow for the use of widely differing numbers of stations through time.

Most rain falls in June or July to September and is associated with the seasonal movement of the Intertropical Convergence Zone (ITCZ). Kidson¹², Lamb³ and Newell and Kidson⁴ suggested that a contributory reason for drier Sahel summers may be a weaker-than-average convergence of the mean moisture flux into West Africa. Newell and Kidson⁴ found that the lower tropospheric south-westerly winds were shallower and weaker in the dry years, although near the surface they tended to penetrate almost as far north as in most wet years. Variations in the northward penetration of the ITCZ cannot be completely neglected but are not thought to be a major determinant of Sahel rainfall, as was once suggested by Winstanley¹³, but afterwards challenged¹⁴.

Based on a hemispheric general circulation model, it was suggested in 1974 that cooler than normal SSTs in the tropical North Atlantic in summer might reduce west Sahel rainfall (A. Gilchrist, unpublished data). Lamb⁶, Lough¹⁵ and Hastenrath⁵ have demonstrated from observations that large-scale SST anomalies in the tropical North and South Atlantic have probably modulated Sahel rainfall in some recent years, although Lough could find no such links prior to 1940. For a different season, Moura and Shukla¹⁶ have used a numerical model to substantiate the influence of tropical Atlantic SST anomalies on rainfall in the Brazilian Nordeste.

Using more complete SST data than previously available, Folland and coworkers^{1,2} provide hints that the present drought epoch might be linked with global scale changes in SST. After 1960–65, marked warming of the Southern Hemisphere and simultaneous cooling of the Northern Hemisphere oceans is observed; the latter reversed before 1980, since when both hemispheres have warmed. However, the North Indian Ocean has tended to warm nearly in phase with the Southern Hemisphere, with strong cooling in the North Pacific and North Atlantic until recently. Figure 1*b* shows annual July–September SST anomalies

for 1901–85 (from a 1951–80 average) for the Southern Hemisphere and North Indian Ocean together (SHNI), minus the remainder of the Northern Hemisphere (RNH) and for the South minus the North Atlantic Ocean. These regional and worldwide trends have been noted and their reality assessed by many authors^{17–23}, using less complete and less well quality-controlled data (except for some North Pacific studies). Problems of separating climatic signals from instrumental biases in marine data have recently been stressed^{24,25}. Strongest support for Fig. 1*b* comes from the pattern of the second empirical orthogonal function of worldwide SST anomalies calculated for all seasons for 1964–79 by Hsiung and Newell²⁰.

The data in Fig. 1 indicate a marked tendency for a relatively warm (cold) SHNI to be associated with drier (wetter) epochs or (sometimes) individual years in the Sahel. The Atlantic curve gives a less clear association (Table 1); thus, any forcing of Sahel rainfall by large-scale SST anomalies may be on a worldwide rather than an Atlantic scale. Rind and Rossow²⁶ have recently drawn attention to the likely sensitivity of the Hadley circulations to very-large-scale changes in SST.

Figure 2*a* shows a worldwide SST difference field for July to September on a 5° × 5° spatial scale between the five driest and

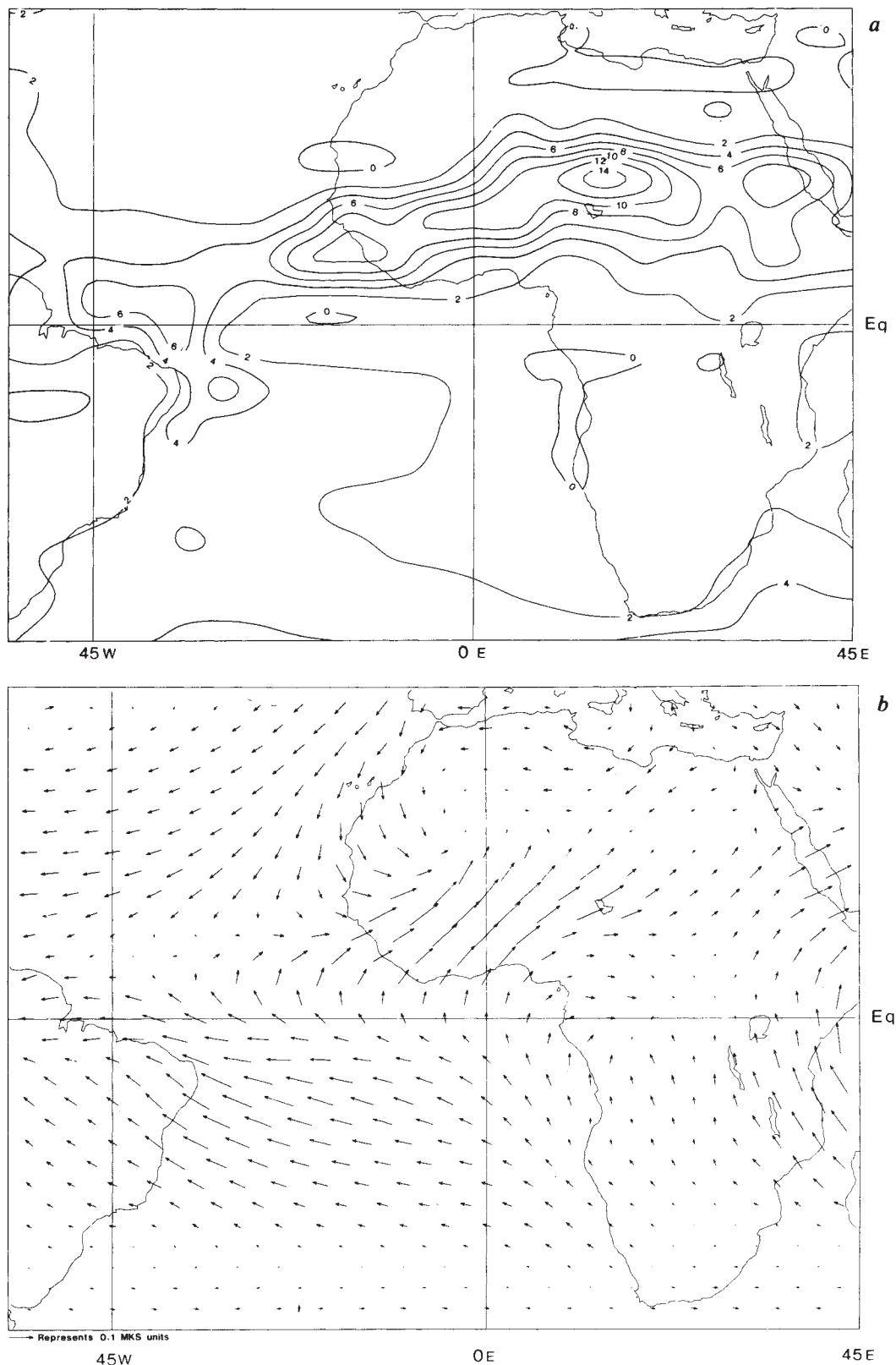
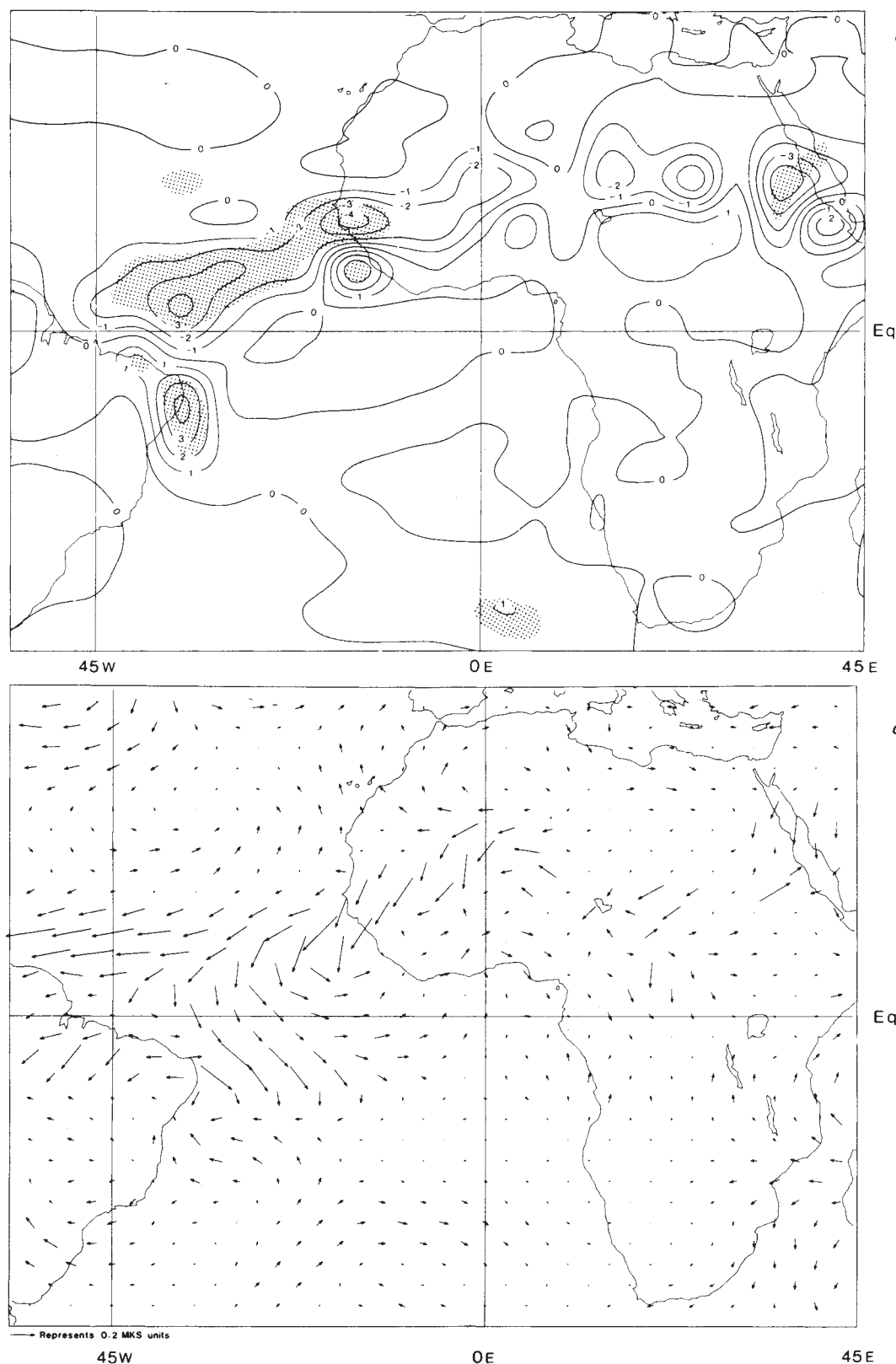


Fig. 3 *a*, Mean rainfall (millimetres per day) in a 180-day control run. *b*, Average 950-mbar steady moisture flux in a 180-day control run. *c*, 180-Day mean rainfall (millimetres per day); anomaly integration minus control. *d*, 180-Day mean 950-mbar steady moisture flux; anomaly integration minus control. The scale of the arrows shown in panel *d* is 5 times that in *b*.

five wettest rainfall seasons in the Sahel since 1945. Differences significant at the 90% level cover 24% of the analysed area, and the pattern is remarkably strong. Figure 2*a* shows that the whole Indian Ocean, South Atlantic and south-east Pacific are warm, while the North Atlantic, North Pacific and Mediterranean are cold. This SST difference pattern reflects a shift in areas of warmest tropical ocean; in particular, areas most subject to deep convection²⁷, that is warmer than 26°C, are reduced

and displaced slightly southward in the Atlantic but considerably increased in size in the western Indian Ocean. The warmth of the tropical south-east and central Pacific in very dry Sahel years reflects the influence of the 1972-73 and 1982-83 El Niño events. A worldwide SST difference pattern corresponding to the eight driest minus the eight wettest Sahel seasons before 1940 gives less clear warmth in the Indian and South Atlantic oceans (except the Gulf of Guinea) in the dry years, although the North



Pacific, North Atlantic and Mediterranean are still cool. An El Niño signature is also absent. Despite the poorer SST and rainfall data prior to 1945, the differences in the patterns may be partly genuine; the driest Sahel years before 1945 tended to accompany relatively wet conditions in West Africa around 9–11° N (Nicholson's¹⁰ Soudano-Guinea series), whereas after 1945 most of western tropical Africa was dry when the Sahel was very dry. Thus the SST anomaly pattern associated with

'Sahel drought' is unlikely to be unique.

If Fig. 2a represents an important SST anomaly pattern that modulates Sahel rainfall on several timescales, it is reasonable to estimate indices of the similarity of each year (in the interval 1901–85) to this pattern (pattern (1)) and to the pattern observed prior to 1940 (pattern (2)) and to correlate both sets of indices year by year with Sahel rainfall. Each index was defined as a modified measure of the covariance between worldwide SST

Table 1 Correlation of July to September SST time series and Sahel rainfall, 1901–84

No.	Ocean SST series	<i>r</i>	<i>S</i>	σ_0 (°C)	<i>c</i>	<i>m</i>
1	Southern minus Northern hemisphere	−0.56	**	0.18	−0.15	−1.69
2	SHNI – RNH	−0.62	***	0.19	−0.15	−1.78
3	South – North Atlantic (0–30° S, 0–30° N)	−0.36		0.32	−0.19	−0.60
4	South – North Atlantic	−0.44	*	0.32	−0.24	−0.75
5	As 4 but with division at 5° N	−0.46	*	0.32	−0.25	−0.77
6	South Atlantic + S. Indian + N. Indian + East Tropical Pacific – N. Atlantic – N. Pacific – Mediterranean	−0.67	***	0.96	−0.23	−0.38
7	As 6 but ocean areas weighted according to size	−0.62	***	0.26	−0.09	−1.26
8	As 6 but with Atlantic split at 5° N	−0.69	***	0.96	−0.23	−0.39

r, Correlation coefficient; *S*, significance of *r* (assuming only 32 degrees of freedom because of lag-autocorrelation). *, = 95%; **, = 99%; ***, = 99.9%; σ_0 , Standard deviation of SST series (°C); *c*, regression constant (units are standardized Sahel rainfall); *m*, regression slope.

anomalies and the SST difference pattern:

$$I_{jm} = \sum_{i=1}^n (SSTA_{ij} \times \Delta T_{im}) - G_j$$

$SSTA_{ij}$ is the SST anomaly in the $5^\circ \times 5^\circ$ area *i*, in year *j*; and ΔT_{im} is the SST difference in Fig. 2a ($m=1$) or that prior to 1940 ($m=2$). G_j is a correction for fluctuations (from the 1951–80 average) in the mean SST measured over the whole pattern, which effectively removes global mean SST trends and instrumental biases. I_{jm} therefore reflects relative SST changes on a large scale. The correlations of I_{jm} with Sahel rainfall for 1901–84 were −0.72 (99.9% significant) for pattern (1) and −0.56 (99% significant) for pattern (2). For 1901–39 they were −0.43 (95% significant; pattern (1)) and −0.76 (99.9% significant; pattern (2)) and for 1946–84 the correlations were −0.85 and −0.48. (The significance of these data is hard to assess because of high autocorrelation between successive years' rainfall).

A stern test of any similarity in the influence of SST patterns (1) and (2) is to calculate the best linear regression equation between I_{jm} and Sahel rainfall S_j for the epoch of the SST difference pattern concerned, and to use the equation to simulate (that is, to predict using simultaneously observed SST anomalies) Sahel rainfall in the independent epoch. Simulated rainfalls for 1901–39 and for 1946–85, derived from I_{j1} and I_{j2} respectively, are plotted with the original values in Fig. 2b. The correlations between observed and simulated rainfall are 0.43 (30 degrees of freedom) and 0.48 (4 d.f.) respectively. These results suggest that a common forcing pattern is present in the summer SST anomaly patterns in both epochs despite a great change in the statistical structure of the Sahel rainfall series, which was dominated by large yearly variability before 1940 but was highly persistent after 1950.

We have carried out numerical experiments with the UK Meteorological Office 11-level AGCM (described by A. Slingo in the unpublished report Met 0 20 DCTN 29, September 1985) to gain more insight into the influence of the SST anomaly difference pattern shown in Fig. 2a. We ran the model in perpetual July mode for 180 days, first with fixed climatological average July SSTs and then with the SST anomaly pattern of Fig. 2a superimposed on the climatological values. Figure 3a shows the model's Sahel rainfall with climatological SSTs; the response of an AGCM to a particular SST anomaly is strongly dependent on its climatology without the anomaly²⁸. Over the western and central Sahel, between about 15° W and 10° E, the model's rainfall climatology is in good agreement with Jaeger's²⁹ observed climatology. East of this region, the model's climatological rainfall extends north and south of the observed limits; the maximum northeast of Lake Chad is excessive. Over the western Sahel, model rainfall exceeds evaporation (not shown) by about 4 mm per day and moisture is supplied to this region by the horizontal moisture flux. The 'steady' component of this flux (that is, mean moisture field $\times$ vector mean wind) is shown in Fig. 3b for the 950 mbar level.

Figure 3c shows the rainfall in the integration with the observed composite SST anomaly, minus the control rainfall. Over the western Sahel, rainfall is decreased on average by about 30% of the model's climatological values. Over the eastern Sahel, there is a smaller but coherent reduction of about 1–2 mm per day, with reductions of up to 50% over the mountains of south-east Sudan and north-east Ethiopia. Figure 3d shows that the model's steady 950 mbar flux of moisture from the South Atlantic into the western Sahel is decreased by about 20–50%. This experiment therefore shows that the observed SST difference pattern of Fig. 2a can provide a coherent change in Sahel rainfall, mainly through a reduction in moisture flux convergence, with perhaps a small southward displacement of the ITCZ in the western Sahel. The effect of warmer SSTs in the Southern Hemisphere in increasing low-level saturation mixing ratios may therefore be much less important than the impact of the SST anomalies on the dynamics of the tropical atmosphere. This conclusion is consistent with the reduced divergence at 250 mbar in the model (not shown) over almost the whole Sahel belt, from south of 10° N up to 20° N and from the Atlantic to the Red Sea.

In a companion paper³⁰, the responses of the AGCM to individual composite SST anomalies in the Atlantic, Pacific and Indian oceans are tested individually. Although the Atlantic had the strongest effect on rainfall over the western Sahel, the Pacific anomalies also had a strong effect, and none of the individual ocean anomalies gave as large an effect as the worldwide pattern of Fig. 2a. It is likely that the Pacific influences Sahel rainfall through the generation of anomalous upper tropospheric westerlies across tropical Africa, similar to those from a northern winter El Niño SST anomaly²⁸. The influence of the Indian Ocean in reducing modelled Sahel rainfall appears to be largest over Sudan and Northern Ethiopia. A full understanding of the effects of SST on Sahel rainfall requires consideration of the global oceans.

Received 26 October 1985; accepted 24 February 1986.

1. Folland, C. K., Parker, D. E. & Kates, F. E. *Nature* **310**, 670–673 (1984).
2. Folland, C. K., Parker, D. E. & Newman, M. R. *Proc. 9th Clim. Diags Conf.*, Corvallis, Oregon, 22–26 Oct. 1984, 70–85 (National Oceanic and Atmospheric Administration).
3. Lamb, P. J. *Tellus* **35A**, 198–212 (1983).
4. Newell, R. E. & Kidson, J. W. *J. Clim.* **4**, 27–33 (1984).
5. Hastenrath, S. *Mon. Weath. Rev.* U.S. Dep. Agric. **112**, 1097–1107 (1984).
6. Lamb, P. J. *Tellus* **30**, 240–251 (1978).
7. Cunningham, W. M. & Rowntree, P. R. *Q.J. R. met. Soc.* **112**, (in the press).
8. Charney, J. G., Quirk, W. J., Chow, S. H. & Kornfield, J. *J. Atmos. Sci.* **34**, 1366–1385 (1977).
9. Courel, M. F., Kandel, R. S. & Rasool, S. I. *Nature* **307**, 528–531 (1984).
10. Nicholson, S. E. *Mon. Weath. Rev.* U.S. Dep. Agric. **108**, 473–487 (1980).
11. Nicholson, S. E. *J. Clim. appl. Met.* **24**, 1388–1391 (1985).
12. Kidson, J. W. *Q. J. R. met. Soc.* **103**, 441–446 (1977).
13. Winstanley, D. *Nature* **243**, 464–465 (1973).
14. Miles, M. K. & Folland, C. K. *Nature* **252**, 616 (1974).
15. Lough, J. M. *Clim. Monit.* **9**, 150–157 (1980); *Mon. Weath. Rev.* U.S. Dep. Agric. **114** (in the press).
16. Moura, A. D. & Shukla, J. *J. Atmos. Sci.* **38**, 2653–2675 (1981).
17. Bunker, A. F. *Mon. Weath. Rev.* U.S. Dep. Agric. **108**, 720–732 (1980).
18. Kukla, G. J. *et al.* *Nature* **270**, 573–580 (1977).
19. Newell, R. E. & Hsiung, J. *Riv. Ital. Geofis. Sci. Affini* **5**, 121–125 (1979).

20. Hsiung, J. & Newell, R. E. *J. phys. Oceanogr.* **13**, 1957-1967 (1983).
21. Paltridge, G. & Woodruff, S. *Mon. Weath. Rev. U.S. Dep. Agric.* **112**, 1093-1095 (1984).
22. Douglas, A., Cayan, D. & Namias, J. *Mon. Weath. Rev. U.S. Dep. Agric.* **110**, 1851-1862 (1982).
23. Cadet, D. L. & Diehl, B. C. *Mon. Weath. Rev. U.S. Dep. Agric.* **112**, 1921-1935 (1984).
24. Barnett, T. P. *Mon. Weath. Rev. U.S. Dep. Agric.* **112**, 303-312 (1984).
25. Oort, A. H. & Maher, M. C. in *Coupled Ocean-Atmosphere Models* (ed. Nihoul, J.) 183-198 (Elsevier, Amsterdam, 1985).
26. Rind, D. & Rossow, W. B. *J. atmos. Sci.* **41**, 479-507 (1984).
27. Gadgil, S., Joseph, P. V. & Joshi, N. V. *Nature* **312**, 141-143 (1984).
28. Palmer, T. N. and Mansfield, D. A. *Q. J. R. met. Soc.* (in the press).
29. Jaeger, L. *Berichte Deutscher Wetterd.* Vol. 18, No. 139 (1976).
30. Palmer, T. N. *Nature* (in the press).

Glaciers as indicators of a carbon dioxide warming

J. Oerlemans

Institute of Meteorology and Oceanography, University of Utrecht, Princetonplein 5, Utrecht, The Netherlands

During the past 150 years, mountain glaciers have shown a world-wide retreat. It has been argued that this is related to the warming which is predicted to result from increased carbon dioxide levels in the atmosphere; however, this warming has not been detected in a statistically significant way from instrumental records. I demonstrate here that the lower part of a valley glacier is extremely sensitive to a local warming, induced by an increase in the radiation budget. For glaciers covering only a small fraction of a valley, the effect is particularly dramatic. Thus valley glaciers may be extremely vulnerable to the presence of infrared-absorbing gases in the atmosphere, and could therefore be better detectors of a possible carbon dioxide warming than is generally assumed.

Climatic indicators such as tree-ring width, sea level and glacier variations can extend our knowledge of the Earth's climatic history beyond the relatively recent period covered by instrumental records. In Europe, the presence of long valley glaciers in areas that have been populated for hundreds of years provides long records of glacier front variations. Figure 1 shows records of three glaciers in the Alps and one in Norway (Nigardsbreen)¹⁻⁵. The dramatic retreat of these large glaciers since 1850 is generally believed to be part of a worldwide retreat, which contributed significantly to the observed rise in global sea level⁶. It is therefore natural to ask whether our knowledge of historic glacier variations can be of some value with regard to the carbon dioxide issue.

Many factors affect the annual mass balance of a glacier⁷⁻¹⁰; depending on type of glacier and climatic conditions, these include winter accumulation, summer temperature, cloudiness and early summer snowfall. Here I concentrate on melting rates at the lower parts of long glaciers, such as those referred to in Fig. 1. Such glaciers have their snout at altitudes where mean air temperature is far above the melting point in summer, and advective transfer of heat from regions surrounding the glacier is a very important factor. As will be demonstrated, this advective transfer amplifies the sensitivity of melting rates to changes in the radiation budget.

The calculation described below is based on a highly simplified model of the energy budget of a valley-glacier system, in which a daily cycle is obtained for four dependent variables; namely, surface temperature of the glacier (θ_s), surface temperature of the surrounding regions (T_s), mean boundary-layer temperature over the glacier (θ_a) and mean boundary-layer temperature over the surrounding regions (T_a). Here, the surface temperature is the mean temperature of the surface layer involved in the daily cycle, and the boundary-layer temperature is the mean air temperature in the lowest 1-1.5 km of the atmosphere.

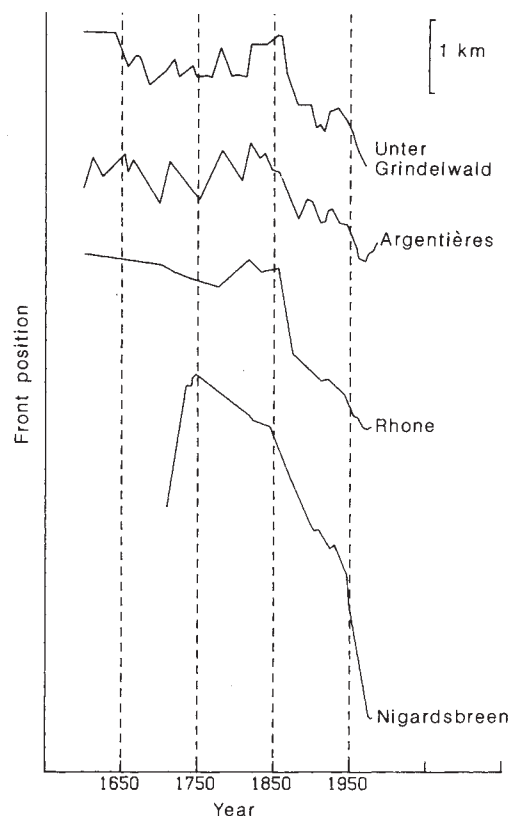


Fig. 1 Front variations of three valley glaciers in the Alps and one in Norway (Nigardsbreen). Data from refs 1-5. The retreat since 1850 is striking.

A diagram of the model is shown in Fig. 2. There are four boxes which interact by radiative processes (single arrows) and by turbulent and advective heat fluxes (double arrows). The surface layers also lose energy to space: a small part of the outgoing infrared flux escapes directly. The background atmosphere has a prescribed characteristic temperature T_b , so the radiation balance of this layer is irrelevant. Absorption of solar radiation in the boundary layer is ignored, so the model is forced by the absorption of solar energy at the surface.

The boundary is assumed to have an effective long-wavelength emissivity μ , while that of the background atmosphere is denoted by ν . The radiation balance of the surface layer (surrounding ground) can then be formulated as:

$$R_s = A Q(t) - \sigma T_s^4 + \mu \sigma T_a^4 + (1 - \mu) \nu \sigma T_b^4, \quad (1)$$

and the expression for the boundary layer is:

$$R_a = \mu \sigma T_s^4 - 2 \mu \sigma T_a^4 + \mu \nu \sigma T_b^4 \quad (2)$$

In equation (1), the first term on the right-hand side represents the absorption of solar radiation at the surface (Q is incident radiation, A effective absorptivity), the second term the upward long-wavelength emission, and the third and fourth terms counter-radiation from the atmosphere. σ is the Stefan-Boltzmann constant. The interpretation of equation (2) is similar, but there is no short-wavelength (solar) contribution. Equivalent equations can be written for the glacier, but the absorptivity will generally be lower.

A few comments on the effective absorptivity A are in order. In the present schematic calculation, effects such as those of cloud cover and atmospheric dust content are all absorbed in this quantity. A simply indicates what fraction of the solar radiation incident at the top of the atmosphere is absorbed at the surface. This implies that in the sensitivity experiments

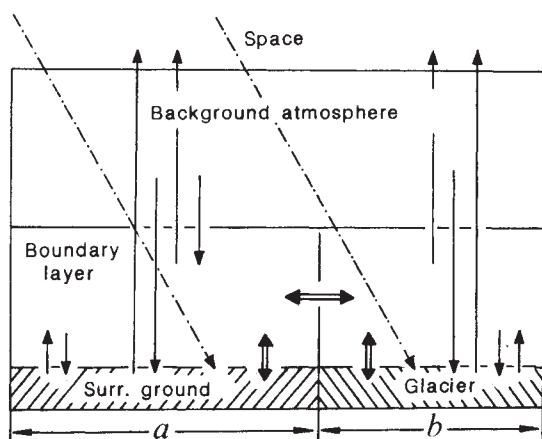


Fig. 2 Schematic representation of the energy fluxes included in the model calculation. The geometry is characterized by length scales a (surrounding ground) and b (glacier). Single arrows, radiative fluxes (solid lines, long-wavelength; broken line, solar). Double arrows, turbulent and advective heat fluxes.

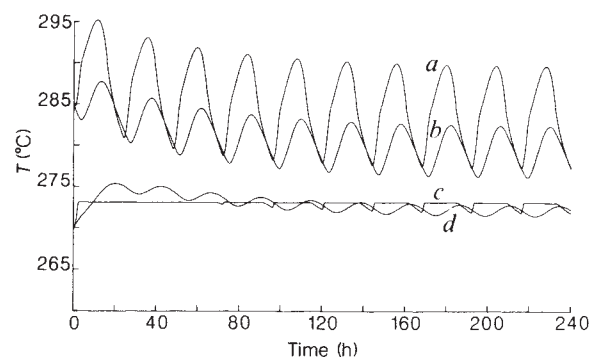


Fig. 3 Standard run showing how the system settles to stationary conditions. The curves are: a , surface temperature of the surrounding ground; b , mean boundary-layer temperature over the surrounding ground; c , surface temperature of the glacier tongue; d , mean boundary-layer temperature over the glacier. For this particular run, model parameters are: $\mu = 0.3$, $\nu = 0.8$, $T_b = 245$ K, $C = 10^6$ J m $^{-2}$ K $^{-1}$, surface-layer heat capacity $= 3 \times 10^5$ J m $^{-2}$ K $^{-1}$, $D = 10$ m 2 s $^{-1}$, $a = 5,000$ m and $b = 1,000$ m. These estimates were derived from standard physical climatology¹³.

discussed below, all factors determining the effective absorptivity remain unchanged.

Turbulent heat fluxes occur between surface layer and boundary layer (vertical fluxes) and advective heat transport between glacier and surroundings in the boundary layer (horizontal fluxes). The vertical fluxes are set proportional to the difference in potential temperature. The exchange coefficient also depends on this difference, because the suppression of heat fluxes in a stably stratified atmosphere is an important feature even for the present schematic model. The dependence on stratification was formulated in agreement with current ideas¹¹. With regard to the daily cycle in fairly stable conditions (which is of interest here), inclusion of an advective heat flux between boundary layer and background atmosphere did not result in significant effects, so it is neglected here.

When considering horizontal energy fluxes, geometric effects come into play more explicitly. If the glacier and the surrounding area are characterized by length scales b and a , respectively, then the quantity $b/(a+b)$ gives the fraction of a valley covered by glacier ice. In a first-order approximation it is reasonable to set the horizontal energy transport proportional to the boundary-layer temperature difference over the glacier and the surrounding regions. The rate equations for T_a and θ_a are then of the form:

$$\frac{dT_a}{dt} = R_a/C - k(T'_a - T_s) - (T_a - \theta_a)D/[a(a+b)] \quad (3)$$

$$\frac{d\theta_a}{dt} = P_a/C - k(\theta'_a - \theta_s) - (\theta_a - T_a)D/[b(a+b)] \quad (4)$$

The prime indicates potential temperature, P_a is the radiation balance of the boundary layer over the glacier, C is the heat capacity of the boundary layer (per unit area), k the vertical and D the horizontal exchange coefficient.

The model is completed by similar rate equations for T_s and θ_s , details of which will not be given here. An additional condition is that θ_s cannot exceed the melting point. When the glacier surface is at the melting point all surplus energy is used for melting.

The system of four coupled equations is easily solved on a computer. An example is shown in Fig. 3; model parameters are given in the legend. The forcing function $Q(t)$ consists of half of a sine function in time (half-period 16 h), with a peak value of 700 W m $^{-2}$. The effective absorptivity is set to 0.5 over the glacier and to 0.75 over the surrounding region.

After about 5 days of integrated time the model temperatures settle to a steady regime. The amplitudes of the daily variation

are quite reasonable, as are the observed phase shifts. T_s and θ_s are almost in phase with $Q(t)$ (not shown). The boundary-layer temperature over the glacier (curve d) shows the largest shift: it reaches a maximum value in the beginning of the evening. This result is difficult to verify from existing observations, but it is probably correct. The amount of melt water produced in the last cycle of this run corresponds to 0.0622 m ice per day.

Calculated melt rates are particularly sensitive to the atmospheric background temperature (T_b) and to the horizontal energy fluxes. Changing T_b by 5 K leads to a change of 2 cm in daily melt. Reducing D by 50% causes a 35% decrease in daily melt, doubling D a 30% increase. This demonstrates that much of the energy used for melting is imported from the region surrounding the glacier tongue.

A series of model runs was carried out to investigate the sensitivity of the melting rates to changes in the radiation balance. One can think of sophisticated ways to mimic the effect of an increasing carbon dioxide concentration in the atmosphere, but here I simply assume that the radiation balance of the surface is perturbed by a constant value of 6 W m $^{-2}$. Such a value is mentioned in the literature as characteristic of a doubling of the carbon dioxide concentration¹². For each of several sets of model parameters, the following runs were carried out: (1) radiation perturbation over the glacier only, (2) radiation

Table 1 Melt rates (mm ice per day) obtained in sensitivity tests

Run	(a, b) (m)				
	(5,000, 5,000)	(5,000, 2,000)	(5,000, 1,000)	(5,000, 500)	(1,000, 200)
Standard	5.45	23.1	62.2	117	134
(1)	6.83	24.8	63.6	118	135
(2)	7.31	27.0	70.6	130	148
(3)	8.56	30.1	76.9	139	158

The standard run corresponds to the case shown in Fig. 3, for various values of the length scales a and b . Run (1), the radiation input to the glacier surface is increased by 6 W m $^{-2}$; run (2), the radiation input to the entire region is increased by this amount; run (3), as for run (2), with an additional increase of 1.5 K in the atmospheric background temperature T_b .

perturbation over the entire region, and (3) as in the previous experiment with an additional increase in T_b of 1.5 K.

The results are shown in Table 1. Those in the top row correspond to the standard case of Fig. 3. If the glacier receives an additional 6 W m $^{-2}$, this leads to an increase in daily melt of 1.4 mm ice. If the entire region has an increased radiation balance, this value increases to 7.1 mm. An additional increase of 1.5 K in T_b would bring the total increase of melt to 14.7 mm.

The effect on melting rates of increasing energy transport to the glacier is largest when the glacier occupies a relatively small area, as is the case for a swiftly retreating glacier tongue.

This sensitivity experiment clearly demonstrates that a glacier can make use of surplus radiation in a valley in a very efficient way. The fact that the surface temperature of the glacier cannot rise above the melting point is crucial. Any increase in radiation thus causes an increased temperature difference between glacier and surrounding regions and thus a larger horizontal energy transfer. This 'oasis effect' disappears when a region becomes completely glaciated.

The glacier may respond in a similar fashion to changes in the background climatic state, which is represented in the present simple model by T_b . The third run, in which an increase in T_b is included, was carried out with this in mind. Even when this effect is not included, melting rates on a small glacier tongue may increase by as much as 15% for a 6 W m^{-2} increase in the radiation input. Through the mechanism discussed here, the glacier retreat establishes a positive feedback. A retreat of the glacier front is associated with decreasing ice thickness and hence with decreasing width of the glacier. In terms of the present model, b thus increases and a further change in melt rate will result. This process can be particularly effective when bare rock is involved.

In this simple calculation I have considered only one of many processes which determine the response of mountain glaciers to rising carbon dioxide levels in the atmosphere. Changes in atmospheric circulation will lead to changes in the amount of cloud cover and precipitation (snow cover in spring could be

particularly important¹⁴). Nevertheless, the direct effect of a change in radiation on melting rates at the glacier tongue would seem to be most important.

Given that the timescale of front variations associated with changes in melt rates at the tongue is very short (of the order of decades), the observed glacial retreat can well be explained by the increasing carbon dioxide level in the atmosphere. Due to the oasis effect, valley glaciers can be considered as very good indicators of climatic change induced by small shifts in the long-term radiation balance.

This research is sponsored by the Ministry of Housing, Physical Planning and Environment (The Netherlands) under contract 611003.01.

Received 5 November 1985; accepted 12 February 1986.

1. Messerli, B., Messerli, P., Pfister, C. & Zumbühl, H. J. *Arct. alp. Res.* **10**, 247-260 (1978).
2. Vivian, R. *Les Glaciers des Alpes Occidentales* (Allier, Grenoble, 1975).
3. Messerli, B. *et al. Z. Gletscher. Glazialgeol.* **9**, 3-110 (1975).
4. International Association of Scientific Hydrology *Fluctuations of Glaciers* Vols 1-3 (UNESCO, Paris, 1967, 1973, 1977).
5. Ostrem, G., Liestol, O. & Wold, B. *Norsk geogr. Tidsskr.* **30**, 187-209 (1977).
6. Meier, M. F. *Science* **226**, 1418-1421 (1984).
7. Martin, S. Z. *Gletscher. Glazialgeol.* **13**, 127-153 (1977).
8. Young, G. J. Z. *Gletscher. Glazialgeol.* **13**, 111-125 (1977).
9. Reynand, L. in *Variations in the Global Water Budget* (eds?) 197-205 (Reidel, Dordrecht, 1983).
10. Kuhn, M. in *Sea Level, Ice and Climatic Change* (ed. Allison, I.) 3-20 (International Association of Scientific Hydrology, Paris, 1979).
11. Louis, J. F. *European Centre for Medium Range Weather Forecasts intern. Rep. no. 4* (Reading, 1977).
12. Ramanathan, V. *J. atmos. Sci.* **38**, 918-930 (1981).
13. Sellers, W. D. *Physical Climatology* (University of Chicago Press, 1965).
14. Kukla, G. in *Sea Level, Ice and Climatic Change* (ed. Allison, I.) International Association of Scientific Hydrology, 79-108 (UNESCO, Paris, 1979).

Iron-silica crystallite nucleation by bacteria in a geothermal sediment

F. G. Ferris*, T. J. Beveridge†‡ & W. S. Fyfe*

* Department of Geology, Faculty of Science, University of Western Ontario, London, Ontario N6A 5B7, Canada

† Department of Microbiology, College of Biological Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada

The recognition of microfossils in ancient sedimentary rocks has confirmed that microorganisms which strongly resemble present-day prokaryotes existed at least 3×10^9 years ago¹. Identification of these structures is a difficult task, complicated not only by the simple morphology of microorganisms, but also by our limited understanding of the processes leading to the mineralization of individual cells and their component parts²⁻⁴. We have investigated these processes in a natural setting, by examining sediment samples from an acidic hot spring in Yellowstone National Park. Electron microscopy and energy-dispersive X-ray spectroscopy of thin-sectioned specimens revealed remains of bacteria inside accumulations of iron and silica. Bacterial-like forms in successive stages of mineral preservation by silica were also found.

Recently, Beveridge *et al.*² simulated low-temperature ($\sim 100^\circ\text{C}$) diagenesis of organic-rich sediments by mixing metal-loaded bacteria with a synthetic sediment of quartz and calcite. The sediment was monitored during the experiment by electron microscopy and energy-dispersive X-ray spectroscopy. This work established that bacteria, particularly their cell walls, bind metallic ions tenaciously during early diagenesis. As the cells were aged, elements from the synthetic sediment were mobilized, and complexed with the metals originally bound by the bacteria. In this way, the cells served as distinct nuclei for the formation of mineral phases which are sometimes found in metalliferous sedimentary rocks.

For the present work, we collected sediment from Cyanidium Creek, an acidic stream (pH 3.1) fed by three hot springs in the south central portion of the Norris Geyser Basin, Yellowstone National Park (USA). The water that flows into the creek contains high levels of dissolved metals, and produces surficial deposits of clay and siliceous mud⁵. A significant portion of the organic matter in the sediment of Cyanidium Creek is of microbial origin, derived from the eukaryotic alga *Cyanidium caldarium* (the only thermophilic photoautotroph of acid springs) and from thermoacidophilic bacteria⁶. Sediment samples were taken directly from the sediment-water interface at four different sites along the stream; the temperature of the water phase at each of the collection sites was 60, 51, 49 and 46°C , respectively. Immediately after the sediment samples were removed from the water, they were placed into metal-free plastic tubes (5 ml) containing 1 ml of 1% (v/v) aqueous glutaraldehyde, a fixative for electron microscopy. When the samples were returned to the laboratory they were prepared for thin-sectioning by dehydration through an ethanol-propylene oxide series and embedding in epoxy resin (Epon 812). Gold-coloured thin-sections were obtained with a Reichert Ultracut E ultramicrotome. Electron microscopy was done at 60 keV using a Philips EM400T

Table 1 Weight percentage of elements in crystallites encrusting bacteria in sediment from Cyanidium Creek

Element	Wt%*
Si	55.5
Fe	19.7
Cl	12.8
As†	9.7
S†	1.3
Al†	0.5
K†	0.5

* Figures represent an average of values obtained from energy-dispersive X-ray spectroscopy point analyses of 15 different bacterial cells.

† These elements were detected in the crystallite matrix of less than half of the bacteria examined.

‡ To whom correspondence should be addressed.

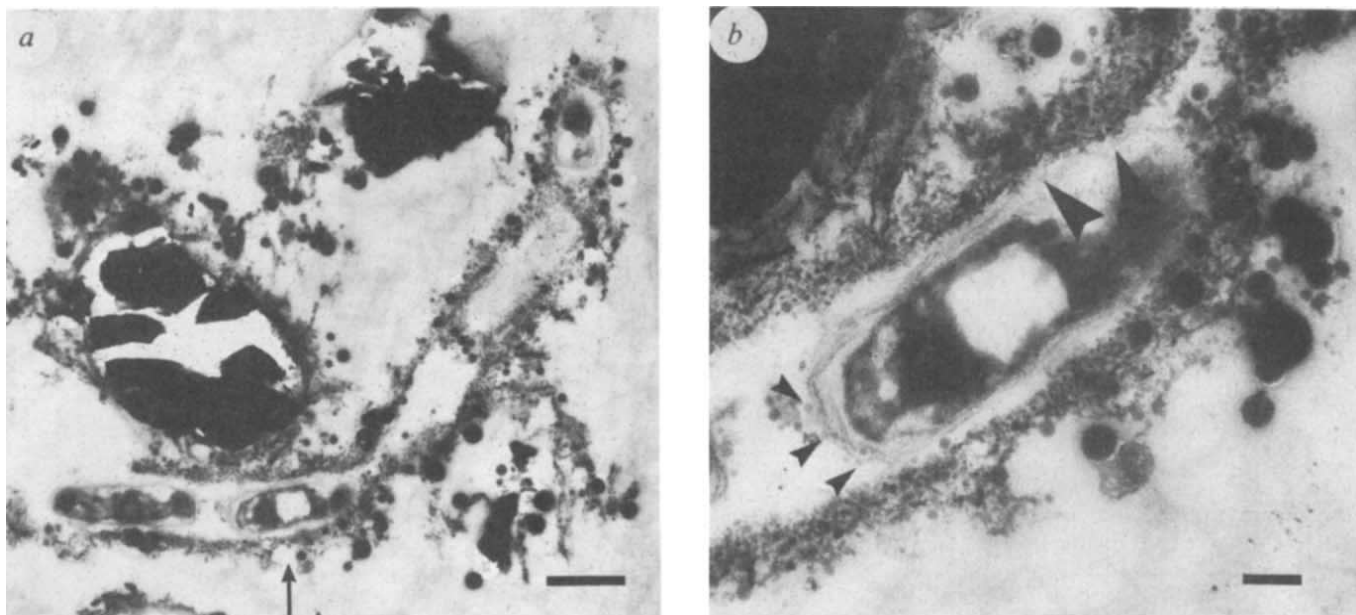


Fig. 1 Thin-section electron micrographs showing a chain of bacteria seen in a sediment sample collected from Cyanidium Creek. The section was stained with uranyl acetate and lead citrate to enhance the contrast of the cells before imaging in the electron microscope. *a*, An electron-opaque matrix of granular and spheroidal crystallites has been formed over the length of the chain. The cells at the ends of the chain are intact, whereas the centre two, although recognizable, have autolysed (scale bar, 500 nm). *b*, Detail of area marked with an arrow in *a*. At this magnification, the fine structure of the matrix is more easily discerned. Crystallites are evident both on (large arrows) and in (small arrows) the cell wall (scale bar, 100 nm).

equipped with an EDAX energy-dispersive X-ray spectrometer.

A chain of rod-shaped bacteria, which were prominent in all sediment samples, is shown in Fig. 1. Cytoplasmic material, which was made visible by heavy-metal-contrasting with uranyl acetate and lead citrate, is evident only in the cells at the ends of the chain. Presumably, the two cells in the middle had autolysed before the sediment was chemically fixed. The entire length of the chain is encrusted in an electron-opaque matrix consisting of granular and spheroidal crystallites. This material was also apparent in unstained thin-sections, and generated electron diffraction patterns indicative of a low-order or near-amorphous state. These crystallites, which are more easily discerned in the portion of the chain shown at a higher magnification (Fig. 1*b*), have grown on, and in, the cell walls of the bacteria. This observation parallels the results of earlier efforts to simulate sedimentary diagenesis in the laboratory².

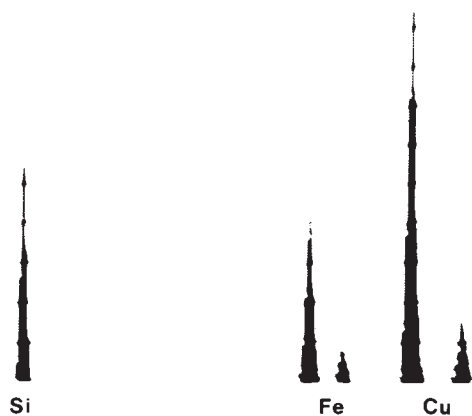


Fig. 2 Energy-dispersive X-ray spectrum generated by the crystallite matrix encrusting a bacterium. The Si (K) and Fe (K α and β) peaks are from the crystallites; Cu (K α and β) peaks are from the grid.

To ascertain the elemental composition of the crystallite mosaic, several point analyses were made using energy-dispersive X-ray spectroscopy. The electron-probe spot size was constant, at 100 nm. The results are summarized in Table 1, and a typical spectrum is shown in Fig. 2. The most abundant element was silicon, suggesting that the bacterial cells promoted the deposition of dissolved silica by serving as sites for crystallite nucleation. Certainly, the small spheres present in the matrix encrusting the bacteria bear a strong resemblance to the silica spheroids which form during the hydrothermal crystallization of amorphous silica gels⁷. We believe that the iron, and other trace-metal elements, were bound as soluble cationic species from the aqueous phase by the constituent homo- and heteropolymers of the bacteria⁸. This was suggested by the strong iron signals detected by energy-dispersive X-ray spectroscopy of cells in which silica deposition was not evident. After being complexed into the sediment by the cell, the metals would either be recycled into the overlying water, or be transformed and immobilized into a mineral phase. The latter process can proceed spontaneously with metallic cations, such as ferric iron, which exhibit unstable aqueous chemistries. Indeed, results from metal binding studies with isolated bacterial cell walls indicate that the removal of metallic ions from solution by binding to bacterial components is often sufficient to induce a transformation to an insoluble hydroxide form (for example, [Fe(OH)₃]_n) (refs 9, 10). We believe that this has occurred here, but it is difficult to illustrate due to the extensive silica deposition in and on the cells.

As yet, the highest degree of preservation exhibited among microfossils appears to have occurred through early diagenetic premineralization by silica¹. Oehler and Schopf¹¹ have shown in experiments designed to simulate these events that microbial cells suffer a minimum of structural damage under conditions in which silica crystallites form slowly. Their best results were obtained in experimental runs lasting up to four weeks at a temperature of 150 °C and pressures of 2–3 kbar. At the lower temperatures and atmospheric pressure which prevail at Cyanidium Creek, silica crystallite nucleation and growth would proceed extremely slowly¹¹. We can therefore assume that the

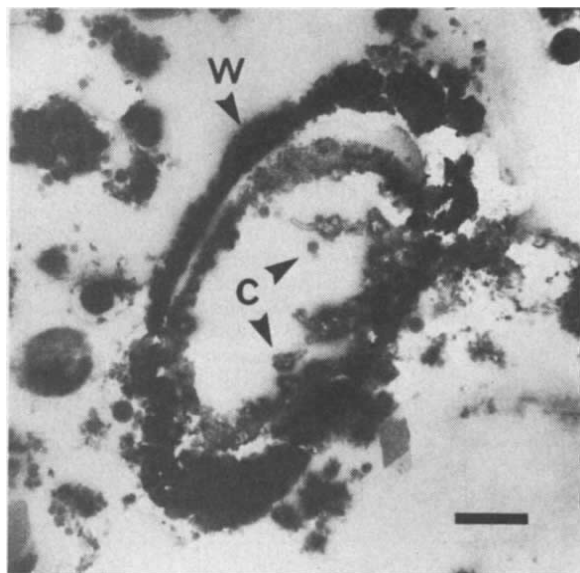


Fig. 3 The mineralized remains of a bacterium after what we presume has been an extended period of iron-silica crystallite growth and compaction. Both the cell wall (W) and cytoplasm (C) have become mineralized (scale bar, 200 nm).

cells shown in Fig. 1 are at least several months old. After what appears to have been an extended episode of silica crystallite formation and compaction, an interlocking mosaic is produced that forms a recognizable facsimile of a bacterium (Fig. 3). The relative abundances of silicon and iron in these compacted structures were similar to those found in the crystallite matrix surrounding cells in earlier phases of silica premineralization. Thus, it is reasonable to expect that during progressive diagenesis in an oxic environment, microdomains of iron oxides, carbonates and silicates would develop from immobilized hydroxide precursors⁴. Conversely, solubilization of iron would be favoured in an anoxic environment.

When microbial cells reach the later stages of mineral preservation, their survival as microfossils depends entirely on whether they become infiltrated and embedded in a supporting

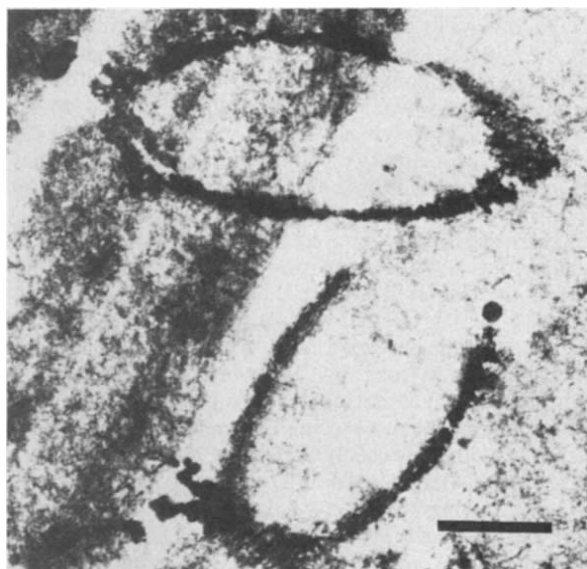


Fig. 4 Completely mineralized bacteria which have become embedded in a fibrous silica matrix (scale bar, 300 nm).

matrix¹¹. For example, the inorganic remains of the bacterium shown in Fig. 3 have already started to fragment. Although this cell would probably not have survived the test of time, the fragmented mosaic of iron-silica crystallites would contribute to a metalliferous sediment, and are probably analogous to the organo-mineralic inclusions observed in the centres of some quartz microspheres in natural cherts⁷. We were fortunate to observe, in one sample, intact preserved bacteria which have become thoroughly embedded in a fibrous silica matrix (Fig. 4). Thus, perhaps some of the microorganisms that are premineralized by silica in hydrogeothermal sediments do survive as structurally intact microfossils. Certainly, the recent discovery¹² of bacterial-like particles in fluid inclusions in quartz crystals collected from the lower geyser basin of Yellowstone National Park supports this point of view.

This work was supported by a Canadian NSERC operating grant to T.J.B. and an Ontario Geoscience Research Program grant to W.S.F. The Philips EM400T was purchased with funds from NSERC and is maintained by funds from the University of Guelph, and user fees. The authors are indebted to R. Humphrey for technical assistance.

Received 28 October 1985; accepted 11 February 1986.

1. Hofmann, H. J. & Schopf, J. W. in *Earth's Earliest Biosphere. Its Origin and Evolution* (ed. Schopf, J. W.) 321-360 (Princeton University Press, 1983).
2. Beveridge, T. J., Meloche, J. D., Fyfe, W. S. & Murray, R. G. E. *Appl. envir. Microbiol.* **45**, 1094-1108 (1983).
3. Trudinger, P. A. & Mendelsohn, F. in *Stromatolites* (ed. Walter, M. R.) 663-672 (Elsevier, New York, 1976).
4. Walker, J. C. G. et al. in *Earth's Earliest Biosphere. Its Origin and Evolution* (ed. Schopf, J. W.) 187-213 (Princeton University Press, 1983).
5. Walker, M. R. in *Stromatolites* (ed. Walter, M. R.) 489-498 (Elsevier, New York, 1976).
6. Brock, T. D. *Thermophilic Microorganisms and Life at High Temperatures* (Springer, New York, 1978).
7. Oehler, J. H. *Geol. Soc. America Bull.* **87**, 1143-1152 (1976).
8. Beveridge, T. J. in *Current Perspectives in Microbial Ecology* (eds Klug, M. J. & Reddy, C. A.) 601-607 (American Society for Microbiology, Washington, 1984).
9. Ferris, F. G. & Beveridge, T. J. *FEMS Microbiol. Lett.* **24**, 43-46 (1984).
10. Beveridge, T. J. & Murray, R. G. E. *J. Bacteriol.* **141**, 876-887 (1980).
11. Oehler, J. H. & Schopf, J. W. *Science* **1229**-1231 (1971).
12. Bargar, K. E., Fournier, R. O. & Theodore, T. G. *Geology* **13**, 483-486 (1985).

Fossil bacterial magnetite in deep-sea sediments from the South Atlantic Ocean

Nikolai Petersen & Tilo von Dobeneck

Institut für Allgemeine und Angewandte Geophysik,
Ludwig-Maximilians-Universität München, Theresienstrasse 41,
8000 München 2, FRG

Hojatollah Vali

Lehrstuhl für Angewandte Mineralogie und Geochemie, Technische
Universität München, Lichtenbergstrasse 4, 8046 Garching, FRG

The magnetostratigraphy of deep-sea sediment cores has proven to be a powerful and high-resolution method of dating and comparing sedimentary records. In contrast to this widespread application, relatively few attempts have been made to analyse and identify the mineralogy, grain size and origin of the carrier of the remanent magnetization^{1,2}. It is widely assumed that dominant ferrimagnetic minerals in deep-sea sediments are of lithogenic origin; that is, that they derive from oceanic or continental detritus, volcanic ashfalls, micrometeorites, hydrothermal precipitation or chemical mineralization processes. A very different view was taken by Kirschvink & Lowenstam³, who suggested that fossil remnants of magnetite-mineralizing organisms, in particular magnetotactic bacteria, might play an important part in the magnetization of sediments deposited in certain detritus-poor aquatic systems. We

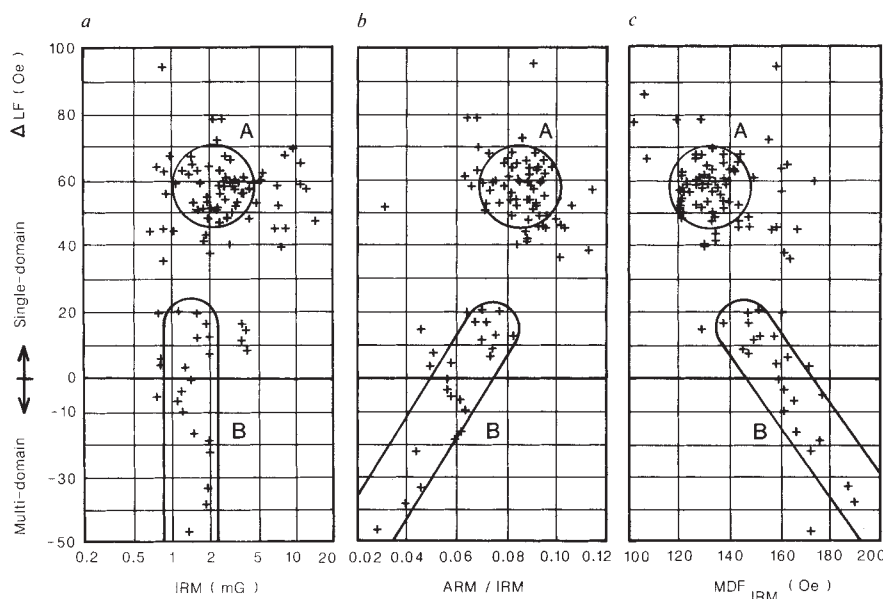


Fig. 1 Functional dependence of Lowrie-Fuller test parameter ΔLF ($\Delta LF = \text{MDF}_{\text{ARM}} - \text{MDF}_{\text{IRM}}$) on *a*, IRM; *b*, ARM/IRM ratio and *c*, MDF_{IRM} . All measurements were performed on whole-rock samples from DSDP sediment cores (Sites 519, 521, 522 and 523; middle Eocene to Quaternary). The IRMs were generated in a magnetic field of 1 kOe, the ARMs in a 0.5-Oe direct field and 1-kOe-peak alternating field. Two sets of data points can be distinguished: set A represents predominantly single-domain particles of presumably biogenic origin; set B also contains coarser magnetic grains of lithogenic origin.

have used rock-magnetic diagnostic methods to characterize the magnetic phases in deep-sea sediments from the South Atlantic; these phases were then extracted and studied with the electron microscope. Our results indicate that fossil bacterial magnetite is the main carrier of the natural remanent magnetization in sediments ranging in age from Quaternary to Eocene. Different types of bacterial magnetosomes were detected, most of them occurring in chains of identical particles, and nearly all of them falling into the theoretically determined⁴ size range of single-domain magnetite.

The samples were taken from DSDP South Atlantic Leg 73 cores (Sites 519, 521, 522 and 523) drilled in 1980 in the Angola Basin⁵, a region comparatively poor in detritus. For the magnetic measurements 110 samples were selected, more or less evenly spaced over the drill cores, and ranging in age from ~5 to 50 Myr. The composition of the samples varies from almost pure nannofossil and foraminiferal ooze (white appearance) to clay-rich material in the Miocene dissolution facies (dark brown colours) with a clay content well over 50% and CaCO_3 less than 10% (ref. 6). Extensive palaeomagnetic measurements have been carried out on these cores in conjunction with nannofossil and foraminiferal analyses, in order to establish a combined magneto- and biostratigraphy^{7,8}. Their rock magnetic properties were studied by Tucker and Tauxe^{9,10}. The aim of our rock-magnetic measurements was to identify the mineralogical nature of the magnetic phases in the sediments (using strong-field thermomagnetic curves, whole-rock hysteresis loops and strong-field IRM (isothermal remanent magnetization) acquisition curves) and to analyse the magnetic domain state and magnetic grain size using various magneto-granulometric tests¹¹⁻¹⁵.

Whereas in mid-ocean-ridge basalts (MORB) the predominant magnetic mineral is titanomagnetite of an average primary composition $\text{Fe}_{2.4}\text{Ti}_{0.6}\text{O}_4$ and Curie temperature $T_C = 150-400^\circ\text{C}$ (depending on the degree of ocean-floor low-temperature oxidation¹⁶), the predominant magnetic phase in the studied sediments has $T_C = 570-600^\circ\text{C}$, suggesting that magnetite (Fe_3O_4) is the main magnetic carrier. In addition, a much less clearly developed T_C of 200–300 °C is often detected in the sediments. We will argue below that this second Curie point is due to the presence of variable amounts of titanomagnetite from altered seafloor basalts. The predominance of magnetite is also reflected in the hysteresis loops, which display saturation at magnetic field strengths below 1 kOe, and in IRM acquisition curves, which show acquisition of 90–95% of the 10-kOe IRM at 1 kOe. The saturation magnetization varied from 5–20 mG

(light sediment) to 100 mG (dark brown sediment), which is equivalent to a theoretical magnetite content of ~0.05–1 mg cm⁻³.

In the isothermal version of the Lowrie-Fuller test¹¹, the anhysteretic remanent magnetization (ARM) acquired in a d.c. field of 0.5 Oe (peak a.c. field 1 kOe) was compared with a 1-kOe isothermal remanent magnetization (IRM) in stepwise AF (alternating field) demagnetization. The results for all samples are summarized in Fig. 1. The difference between the median destructive fields (MDF) of ARM and IRM (referred to as ΔLF in Fig. 1) is plotted against various other parameters. It should be stressed that ΔLF cannot unambiguously indicate the magnetic grain size when different magnetic minerals are present in substantial concentrations. We speak therefore of apparent domain-state behaviour rather than of actual domain state and discuss the ΔLF values only in conjunction with other parameters.

Figure 1a shows that the apparent single-domain (SD) or multi-domain (MD) behaviour is independent of the overall concentration of the magnetic phase (here expressed by the IRM, but equally reflected by plotting ΔLF against the saturation magnetization). The grouping of the data points in Fig. 1a permits a distinction of two magnetically distinct sediment types: type A is clearly SD-dominated and shows extremely little scatter in its intensive parameters. Of the samples studied 40% fall within the circle signifying a ΔLF variation of only ± 12 Oe. This very high homogeneity with absolute MDF_{ARM} values of ~190 Oe is unexpected for sediments which range over most of the Tertiary, with widely varying sedimentation conditions. Type B is characterized by a wide ΔLF distribution and increasing MD behaviour. Sediments of this type are found in ponded local seafloor structures close to the mid-ocean ridge.

In Fig. 1b ΔLF is plotted against the ARM/IRM ratio; the positive correlation observed is in accordance with theory and earlier studies by Cisowski¹². However, the result of Fig. 1c appears paradoxical at first sight: the coercivity—here represented by the MDF_{IRM} value—actually increases with diminishing ΔLF , that is, with supposedly increasing grain size, and is at a minimum for SD particles. We interpret this apparent contradiction in the following way: the type-A phase comprises stress-free, homogeneous and magnetically interacting SD-magnetite particles, whereas in the type-B phase, detrital titanomagnetite grains with higher coercivities due to cation-deficiency-induced internal stress¹⁷ occur in variable amounts together with an omnipresent type-A phase.

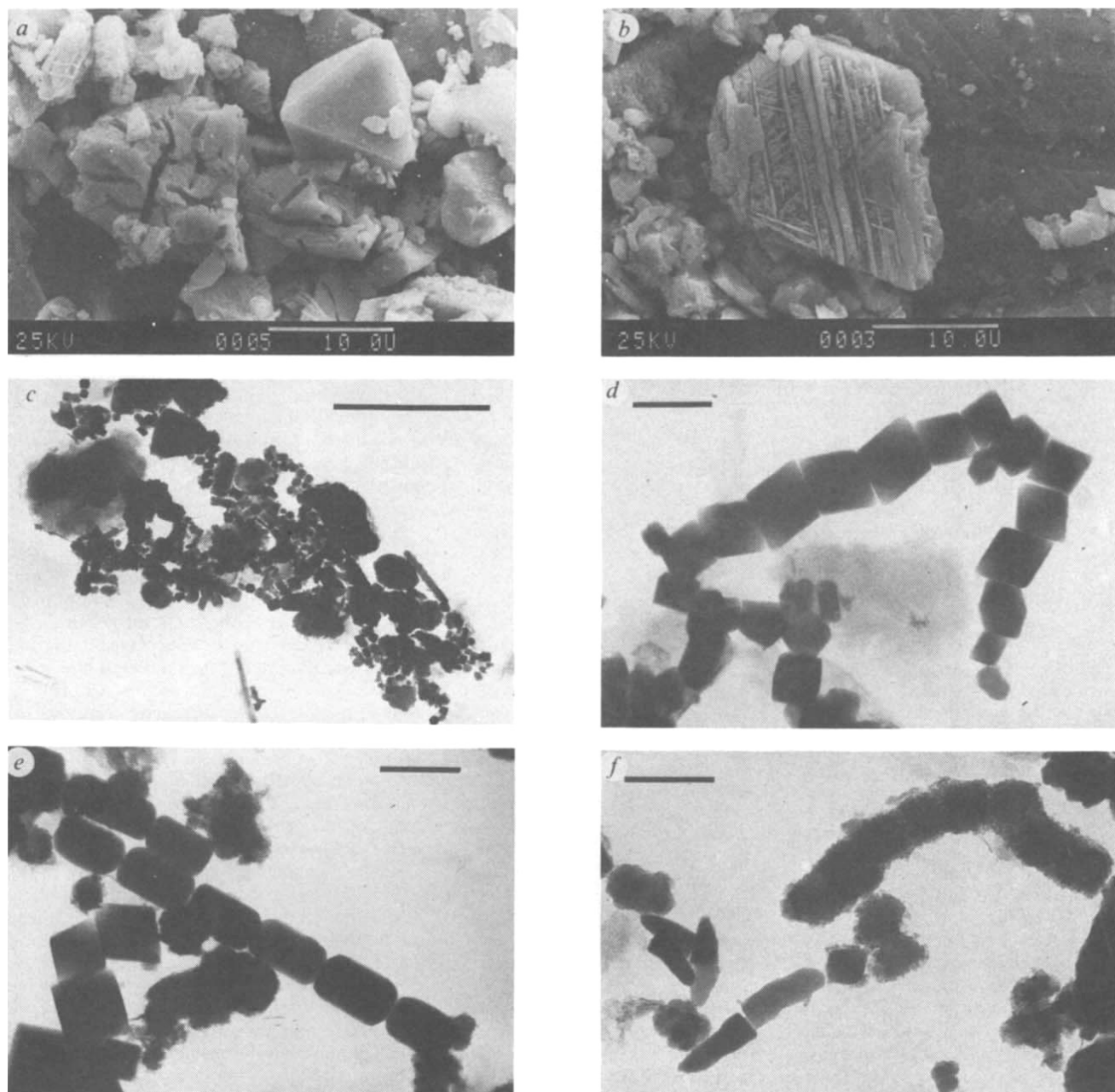


Fig. 2 *a, b*, SEM images of the magnetically extracted coarse fraction of sample 522-15. The separate was etched in hydrofluoric acid to reveal internal grain structures. *a*, Left side, fragmental or skeletal titanomagnetite with abundant irregular cracks, identified as shrinking cracks in MORB titanomagnetites caused by ocean-floor low-temperature oxidation. The iron/titanium ratio corresponds roughly to a composition $\text{Fe}_{2.4}\text{Ti}_{0.6}\text{O}_4$. This group of titanomagnetites presumably derived from altered seafloor basalts, dominates the coarse fraction. Right side, euhedral Ti-Mn-rich spinel, possibly of pyroclastic origin. It is representative of a group of well-developed octahedra containing, in addition to Fe and Ti, various additional cations such as Mg, Al, Mn and Cr in variable proportions. These compositions are typical for titanomagnetites in alkaline basalts. They are quite common in the studied samples and usually show no signs of transport erosion. *b*, Magnetite-ilmenite intergrowths. These typically occur in subaerial basalts and are caused by deuteric high-temperature oxidation of originally homogeneous titanomagnetites. The ilmenite lamellae protrude because the magnetite phase is preferentially dissolved. These grains are less common in the investigated samples. *c-f*, TEM images of the magnetically extracted submicron fraction of various samples from Site 521, cores 3, 6, 12 and 19. *c*, A general view of the submicron fraction. Large fragmental grains are mixed with geometrically well-defined microcrystals of uniform grain size. Chains and clusters of nearly identical octahedra (*d*), parallelepipeds (*e*) and tear-drop-shaped cones (*f*) are common. Some of them are overgrown by smaller, presumably superparamagnetic particles (*f*, right side). Scale bars represent: *a, b*, 10 μm ; *c*, 1 μm ; *d-f*, 0.1 μm .

Interestingly, there is a clear relationship between the colour of the sample (that is, the clay content) and all measured extensive magnetic quantities, such as saturation magnetization, low-field susceptibility and high-field paramagnetic susceptibility (deduced from the linear segment of the hysteresis loop), all these parameters increasing with clay content. This relationship, however, is not observed for the intensive magnetic

parameters related to magnetomineralogy and domain-state behaviour.

Although the rock-magnetic measurements indicate the presence of a predominant, highly homogeneous SD-magnetite phase in most of the samples studied, these measurements cannot reveal whether this magnetic material is of lithogenic or biogenic origin. Kirschvink's^{3,18} fossil magnetosome theory can only be

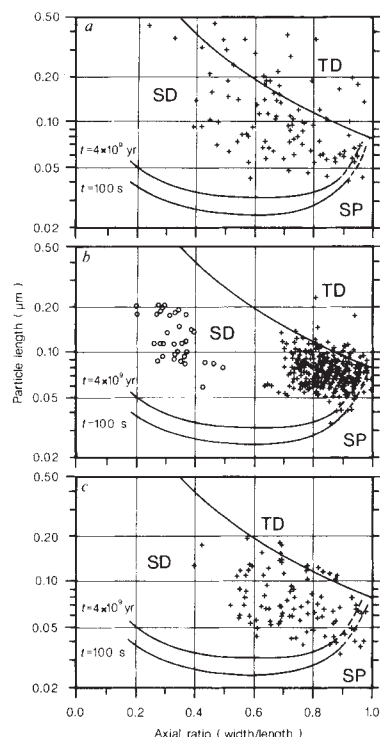


Fig. 3 Grain size distribution of the submicron fraction of the magnetic separates from 521-3, 521-6, 521-12 and 521-19. The length and width of 800 grains were measured from TEM micrographs and plotted as shown. Boundaries between the two-domain (TD), single-domain (SD) and superparamagnetic (SP) stability fields were drawn according to the calculations of Butler and Banerjee⁴. The superparamagnetic threshold is shown for critical relaxation times of 100 s and 4×10^9 yr. *a*, Fragmental and geometrically undefined particles (+); *b*, octahedra (+) and tear-drop shaped cones (O); *c*, parallelepipeds (+). The rod-shaped particles mentioned in the text were not included in this figure.

verified by direct visual observation of the magnetic fraction. Magnetosomes have very distinct shapes¹⁹⁻²¹, and are thus easily distinguishable from lithogenic magnetite.

Magnetite particles in the SD size range (shape-dependent, 300–1,000 Å) are clearly below the optical resolution of a light or scanning electron microscope (SEM) and require transmission electron microscopy (TEM) for direct observation. Because of its low concentration in the sediment, the magnetic phase must be extracted before it can be observed. Customary magnetic extraction techniques are usually unable to collect submicrograin fractions because of physical problems¹, whereas chemical methods²² tend to destroy characteristic grain shapes by partial dissolution²³. To overcome these problems we developed the following non-destructive technique.

The firm, but unsolidified sediment samples were gently disintegrated in a small volume of water and thoroughly dispersed by high-power ultrasonics. A small amount of sodium pyrophosphate was added to prevent coagulation of the suspended clay minerals. After further dilution to a final sediment concentration of ~0.5 vol.% the suspension was allowed to circulate in a special magnetic extraction apparatus. The concept of this system involves two principles which favour the specific extraction of SD-magnetite: one is to provide magnetic field gradients which are as high as possible, while the absolute field intensity is kept relatively low. This results in the attraction predominantly of particles with a permanent magnetic moment. The second is that each grain should have to migrate only a very short distance in the gradient field before it is collected in the magnetic fraction; this supports the aggregation of small, slow-migrating particles.

Thus a cumulative separate carrying 30–40% of the initial sample's low-field susceptibility could be attained in a 1-week cycle. This yield seemed satisfactory to us, as a considerable percentage of the susceptibility of the original material was due to the paramagnetic fraction in the sediment. The separate was then re-suspended and washed several times in a small volume of distilled water to remove clay fractions and other contaminants. Grains larger than ~1 μm were removed by attaching the flat, low-gradient end of a permanent disk magnet to the side of the beaker, allowing the fast-migrating coarser particles to aggregate at the wall, and pouring off the fluid fraction containing the submicron fraction.

Viewing the coarser fraction with a JEOL 30C SEM, we found mainly titanomagnetite grains which fall into three groups according to shape and chemical composition (Fig. 2*a, b*): MORB titanomagnetite fragments or skeletons with signs of ocean-floor low-temperature oxidation, euhedral Fe-Ti-Mg-Al-Mn-Cr spinels with varying composition and, much less abundant, titanomagnetites with ilmenite exsolution lamellae as they typically occur in subaerial lava flows. Pure magnetite particles are very scarce in the coarse fraction and thus cannot represent the predominant magnetic phase indicated by the Curie temperature measurements. No ferrimagnetic sulphides were detected.

The submicron fraction was prepared for TEM observation by simply letting a drop of separate suspension dry off on the carrier grid. Images obtained with a JEOL 100CXTEM at magnifications from $\times 20,000$ to $\times 100,000$ show mostly grain types which are completely different from those observed in the coarse fraction (Fig. 2*c–f*). The dominant portion of the fine fraction consists of geometrically well-defined particles within a very narrow size range (Fig. 2*c*). These microcrystals found in all samples studied had several features that are unknown for lithogenic magnetite. We consider them to be fossil magnetosomes of magnetotactic bacteria which lived in the sediment ooze when it was deposited as long ago as 50 Myr. Our assumption is based on the following observations. Most of the particles are identical in size and shape to magnetosomes in living magnetotactic bacteria¹⁹⁻²¹: octahedra (Fig. 2*d*), slightly rounded parallelepipeds (Fig. 2*e*), and straight or slightly bent cones (Fig. 2*f*). We also observed particles with shapes that have not yet been found in living bacteria, but are probably not of lithogenic origin (such as some with a triangular outline).

Chains or clusters of identical particles were clearly more common than can be accounted for by statistical aggregation processes. This suggests that these configurations are indeed fossil remnants of former bacterial magnetosome chains.

Of 800 submicron grains measured along two perpendicular axes (Fig. 3), all but the randomly shaped, presumably lithogenic fragments (Fig. 3*a*) were within the narrow SD size ranges calculated by Butler and Banerjee⁴. If we accept the view that only SD particles can be used for orientation by a bacterium, this close grouping is a strong argument for the biogenic origin of these particles.

Occasionally we also observed rod-shaped magnetic particles in the submicron fraction, which we suspect to be a new ferrimagnetic mineral forming in iron-rich clay as described by Moskowitz and Hargraves²⁴. Another grain fraction not yet mentioned consists of very fine particles (≤ 100 Å) of unknown composition, which partially or completely overgrow SD grains (Fig. 2*f*), and which are presumed to be superparamagnetic.

We thank L. Bachmann for allowing us to use his electron microscopes and for stimulating discussions, and G. Morteani for encouragement. This research was supported by the Deutsche Forschungsgemeinschaft.

Received 11 October 1985; accepted 5 February 1986.

1. Kent, D. V. & Lowrie, W. *J. geophys. Res.* **79**, 2987–3000 (1974).
2. Løvlie, R., Lowrie, W. & Jacobs, M. *Earth planetary Sci. Lett.* **15**, 157–168 (1971).
3. Kirschvink, J. L. & Lowenstam, H. A. *Earth planet. Sci. Lett.* **44**, 193–204 (1979).

4. Butler, R. F. & Banerjee, S. K. *J. geophys. Res.* **80**, 4049-4058 (1975).
5. Hsü, K. J., LaBrecque, J. & Pisciotto, K. A. *Init. Rep. DSDP 73*, 5-26 (1984).
6. Karpoff, A. M. *Init. Rep. DSDP 73*, 515-535 (1984).
7. Tauxe, L., Tucker, P., Petersen, N. & LaBrecque, J. L. *Init. Rep. DSDP 73*, 609-621 (1984).
8. Poore, R. Z. *et al. Init. Rep. DSDP 73*, 645-655 (1984).
9. Tucker, P. *Init. Rep. DSDP 73*, 663-672 (1984).
10. Tucker, P. & Tauxe, L. *Init. Rep. DSDP 73*, 673-685 (1984).
11. Lowrie, W. & Fuller, M. *J. geophys. Res.* **76**, 6339-6349 (1971).
12. Cisowski, S. *Phys. Earth planet. Inter.* **26**, 56-62 (1981).
13. Day, R., Fuller, M. & Schmidt, V. A. *Phys. Earth planet. Inter.* **13**, 260-267 (1977).
14. King, J., Banerjee, S. K., Marvin, J. & Özdemir, Ö. *Earth planet. Sci. Lett.* **59**, 404-419 (1982).
15. Levi, S. & Merrill, R. T. *J. geophys. Res.* **83**, 309-323 (1978).
16. Bleil, U. & Petersen, N. *Nature* **301**, 384-388 (1982).
17. Appel, E. & Soffel, H. C. *Geophys. Res. Lett.* **3**, 189-192 (1984).
18. Kirschvink, J. L. *Rev. Geophys. Space Phys.* **21**, 672-675 (1983).
19. Blakemore, R. P. & Frankel, R. B. *Scient. Am.* **245**, No. 6, 42-49 (1981).
20. Blakemore, R. P. *A. Rev. Microbiol.* **36**, 217-238 (1982).
21. Towe, K. M. & Moench, T. *Earth planet. Sci. Lett.* **52**, 213-220 (1981).
22. Chang, S. R. & Kirschvink, J. L. in *Magnetite Biomineralization and Magnetoreception in Organisms* (eds Kirschvink, J. L., Jones, D. S. & MacFadden, B. J.) 647-669 (Plenum, New York, in the press).
23. Kirschvink, J. L. & Chang, S. R. *Geology* **12**, 559-562 (1984).
24. Moskowitz, B. M. & Hargraves, R. B. *Science* **225**, 1152-1154 (1984).

Light-induced giant dipoles in simple model compounds for photosynthesis

John M. Warman*, Matthijs P. de Haas*,
M. N. Paddon-Row†, E. Cotsaris†, N. S. Hush‡,
H. Oevering§ & J. W. Verhoeven§

* Interuniversity Reactor Institute, Mekelweg 15, 2600 GA Delft, The Netherlands

† Department of Chemistry, University of New South Wales, PO Box 1, Kensington, New South Wales 2033, Australia

‡ Department of Theoretical Chemistry, University of Sydney, New South Wales 2006, Australia

§ Laboratory for Organic Chemistry, University of Amsterdam, Nieuwe Achtergracht 129, 1018 WS Amsterdam, The Netherlands

The primary steps in photosynthesis involve very rapid (sub-nanosecond) electron transfer between molecular entities that are rigidly embedded within a lipid membrane and separated from each other by well-defined distances on the order of 10 Å. In an attempt to simulate such systems we have studied photon-induced electron transfer within specially synthesized molecular assemblies in which a donor moiety is separated from an electron acceptor by a rigid, saturated hydrocarbon framework of variable length, from 5 to 13 Å. We find charge separation to occur on a sub-nanosecond timescale with close to unit quantum efficiency in all cases. The lifetimes of the resulting charge-transfer states, with dipole moments approaching 70 debye units, can extend to several hundred nanoseconds. Non-conjugated hydrocarbon bridges may be important in determining the rate and direction of electron transfer in photo-excited natural or artificial molecular systems.

Interest in rigid molecular assemblies of donor and acceptor moieties as model systems for photosynthesis has increased dramatically in recent years¹⁻¹⁸. The occurrence of electron transfer following light absorption in such model systems is usually detected by observing the quenching of the local fluorescence of the absorbing chromophore and/or by the appearance of emission or absorption bands characteristic of a charge-transfer state. Until recently, it was not possible to make a direct electrical measurement of the magnitude of the photon-induced charge separation. Here we present quantitative results on light-induced charge separation, obtained by applying the time-resolved microwave conductivity (TRMC) technique¹⁹.

Figure 1 shows the structures of the four compounds investigated; their preparation and purification will be described elsewhere²⁰. They will be referred to in the text as I(4), II(6), III(8) and IV(10). The number in parentheses refers to the number of σ -bonds separating the donor (dimethoxynaphthalene) and acceptor (dicyanoethylene) groups. The cross-bridging of the

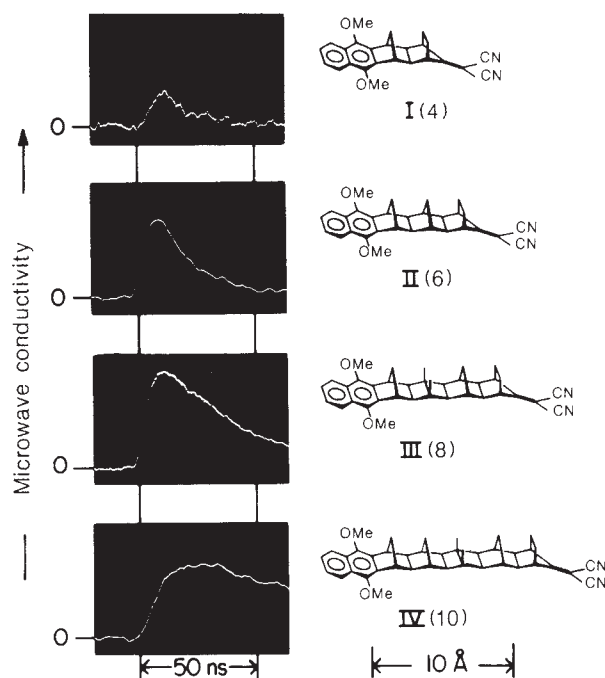


Fig. 1 Microwave conductivity transients resulting from flash photolysis (308 nm, 9.5 mJ cm⁻², 10 ns FWHM) of de-aerated benzene solutions of the compounds shown on the right. The transients are due to rapid electron transfer from the donor to the acceptor in the S₁ state, resulting in charge-transfer states with very large dipole moments. The height of the transients is arbitrary because different vertical sensitivities and solution optical densities were used. The apparent delayed growth for IV(10) results from tailing of the laser pulse and the long transient lifetime.

connecting ring systems ensures complete rigidity of the molecular assemblies.

Excitation was carried out at a wavelength of 308 nm, using a Lambda Physik EMG 500 excimer laser in the single-shot mode, with an integrated intensity per pulse of $\sim 1.5 \times 10^{16}$ photons (10 mJ) per cm² and a pulse width (FWHM) of 10 ns. Only the donor moiety has a significant extinction coefficient at the wavelength used ($\sim 3 \times 10^3$ mol⁻¹ cm⁻¹). The solute concentrations used were within the range 0.1-0.3 mM, and the observed signals remained unchanged after several pulses. The microwave detection system and method of data analysis are described in ref. 19; the time response of detection was 5 ns. The solutions of the compounds in benzene were purged with N₂ to remove oxygen, and with CO₂ to eliminate the possibility of contributions to the conductivity from mobile electrons formed in low-yield secondary photoionization events.

Using this technique, the change in the high-frequency conductivity, $\Delta\sigma$, or dielectric loss, $\Delta\epsilon'' = \Delta\sigma/\omega\epsilon_0$, of the medium on flash photolysis is measured at ~ 10 GHz. The change in conductivity is related to the dipole moment of photoexcited solute molecules by¹⁹

$$\Delta\sigma = \frac{(\epsilon + 2)^2}{27 k_B T} \frac{\Delta p_*^2}{\tau_r} F(\omega\tau_r) N_* \quad (1)$$

where k_B is the Boltzmann constant; ϵ is the relative dielectric constant of the solvent; $\Delta p_*^2 = p_*^2 - p_0^2$, where p_* and p_0 are respectively the dipole moment of the excited state and ground state of the molecule; N_* is the concentration of excited molecules; τ_r is the rotational relaxation time of the dipole and ω is the radian microwave frequency. The function $F(\omega\tau_r)$ is the Debye relaxation term, which for the large solute molecules studied here (with $\omega\tau_r \gg 1$) is equal to unity. Because the donor

and acceptor are not coupled in the ground state, p_0 is expected to be only on the order of 1–2 debye units (D). This is negligible compared with the excited-state dipole moments measured (see below).

Figure 1 shows transient microwave conductivity traces observed for solutions of the compounds in benzene. All four compounds give readily measurable signals, corresponding to large increases of the dipole moments of the solute molecules. Even without further quantitative analysis, therefore, it can be concluded that photon absorption by the donor moiety is followed by rapid electron transfer (Fig. 2), even when donor and acceptor are separated by up to 10 carbon-carbon σ -bonds. This is in agreement with the conclusion already reached regarding the first three compounds, based on quenching studies of the local donor fluorescence and the appearance of redshifted charge-transfer emissions¹ from these compounds in di-*n*-butyl ether. Weak intramolecular charge-transfer emissions in the 500–600-nm region have now been observed for all four compounds in benzene. From the earlier optical measurements¹, the mean time for the charge-separation step, τ_{cs} , was estimated to be less than 20 ps for I(4) and II(6), and 70 ps for the eight- σ -bond molecule, III(8).

The TRMC measurements show that the light-induced charge separation occurs on a sub-nanosecond timescale for all four systems in Fig. 1. As can be seen from the traces in the figure, the lifetime of the charge-transfer state towards recombination, τ_{cr} , increases dramatically as the distance between donor and acceptor is increased. Table 1 lists the lifetimes of the charge-transfer intermediates and their dipole moments, obtained from

Table 1 Lifetimes, rotational relaxation times and dipole moments of the charge-transfer states of the compounds shown in Fig. 1 in benzene solution

Compound	τ_{cr} (ns)	τ_r (ns)	p_* (D)	p_{ee} (D)
I(4)	~1*	0.41	(25)†	24
II(6)	6	0.64	37	37
III(8)	32	0.95	55	49
VI(10)	360‡	1.20	67	61

The values of p_* , calculated assuming unit efficiency for electron transfer, are compared with p_{ee} , the dipole moment expected for complete edge-to-edge charge separation. The values of τ_r were interpolated from data on long-axis-tumbling relaxation times²¹.

* From in-pulse conductivity, based on $p_* = 25$.

† Value determined in cyclohexane.

‡ Corrected for self-quenching.

a quantitative analysis of the data using estimated values of τ_r . The values of τ_r were interpolated from data on long-axis-tumbling relaxation times for rigid rod-shaped molecules determined by both microwave dielectric relaxation measurements and fluorescence depolarization²¹.

The dipole moments estimated from the experimental data are even larger than the values which would be expected for complete charge separation over the edge-to-edge distance between donor and acceptor. It can be concluded, therefore, that electron transfer does in fact occur with close to unit efficiency, and hence that the transfer time is an order of magnitude or more shorter than the ~5-ns lifetime of the S_1 state of the donor (τ_D), even when the transfer distance is as large as 13 Å, as in IV(10).

The dipole moments found in the present study attain values which are much larger than the highest dipole moment ever previously determined (~25 D) for the excited singlet state of a donor-acceptor compound with π -bond coupling between donor and acceptor groups²². The importance of decoupling between donor and acceptor in maximizing the efficiency of conversion of photon energy into electrical potential energy is clearly demonstrated. In view of the large distances involved in

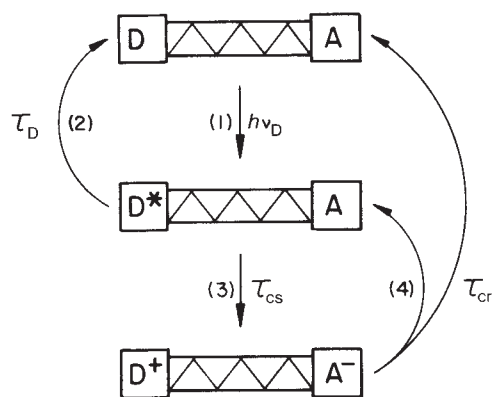


Fig. 2 Schematic representation of the processes occurring on photolysis of the rigid molecular assemblies shown in Fig. 1. (1), Local excitation in the absorption band of the donor moiety, D, by photons of energy $h\nu_D$; (2), local-excited-donor to ground-state relaxation with lifetime τ_D ; (3), electron transfer from local-excited-donor to acceptor moiety, A, in a mean time τ_{cs} ; (4), charge recombination resulting in a lifetime τ_{cr} of the highly dipolar, charge-separated state.

electron transfer in the compounds studied here, the residual weak coupling necessary to initiate electron transfer cannot be attributed to direct spatial overlap between the π -orbitals of the donor and acceptor groups. It must therefore be attributed to through-bond interactions¹, in which the connecting σ -bonded framework acts as a relay.

In its speed, efficiency and magnitude, the photo-induced charge separation achieved in the compounds in Fig. 1 is reminiscent of that occurring in the first two electron-transfer steps of photosynthesis²³. This similarity is perhaps not merely fortuitous, as the recently resolved structure of the photosynthetic reaction centre of *Rhodospseudomonas viridis*²⁴ has shown the isoprenoid phytol side chains of both the special chlorophyll pair and menaquinone to be in contact with one of the two bacteriopheophytins present. The possible role of these chains as relays in mediating the special-pair to pheophytin to quinone electron-transfer steps is clearly worthy of further investigation.

We thank Dr Cyril Varma for the use of flash-photolysis facilities at the University of Leiden, and Petra Weissenborn for assistance. M.N.P.-R., E.C. and N.S.H. thank the Australian Research Grants Scheme for support.

Received 9 September 1985; accepted 12 February 1986.

- Hush, N. S. *et al. Chem. Phys. Lett.* **117**, 8–11 (1985).
- Pasman, P., Koper, N. W. & Verhoeven, J. W. *Recl Trav. chim. Pays-Bas Belg.* **101**, 363–364 (1982).
- Pasman, P., Rob, F. & Verhoeven, J. W. *J. Am. chem. Soc.* **104**, 5127–5133 (1982).
- Mes, G. F., van Ramesdonk, H. J. & Verhoeven, J. W. *J. Am. chem. Soc.* **106**, 1335–1340 (1984).
- Mes, G. F. *et al. J. Am. chem. Soc.* **106**, 6524–6528 (1984).
- Bolton, J. R. *et al. JCS chem. Commun.* 559–560 (1985).
- Stein, C. A. & Taube, H. *J. Am. chem. Soc.* **103**, 693–695 (1981).
- Stein, C. A., Lewis, N. A. & Seitz, G. *J. Am. chem. Soc.* **104**, 2596–2599 (1984).
- Beratan, D. N. & Hopfield, J. J. *J. Am. chem. Soc.* **106**, 1584–1594 (1984).
- Calcaterra, L. T., Closs, G. L. & Miller, J. R. *J. Am. chem. Soc.* **105**, 670–671 (1983).
- Miller, J. R., Calcaterra, L. T. & Closs, G. L. *J. Am. chem. Soc.* **106**, 3047–3049 (1984).
- Overfield, R. E., Scherz, A., Kaufmann, K. J. & Wasielewski, M. R. *J. Am. Chem. Soc.* **105**, 5747–5752 (1983).
- Joran, A. D., Leland, B. A., Geller, G. G., Hopfield, J. J. & Dervan, P. B. *J. Am. chem. Soc.* **106**, 6090–6092 (1984).
- Wasielewski, M. R. & Niemezyk, M. P. *J. Am. chem. Soc.* **106**, 5043–5045 (1984).
- Nocera, D. G., Winkler, J. R., Yocom, K. M., Bordignon, E. & Gray, H. B. *J. Am. chem. Soc.* **106**, 5145–5150 (1984).
- Isied, S. S., Worsila, G. & Atherton, S. J. *J. Am. chem. Soc.* **104**, 7659–7661 (1982).
- Simolo, K. P., McLendon, G. L., Mavic, M. R. & Mavic, A. G. *J. Am. chem. Soc.* **106**, 5012–5013 (1984).
- McGourty, J. M., Blough, N. V. & Hoffman, B. M. *J. Am. chem. Soc.* **105**, 4470–4472 (1983).
- de Haas, M. P. & Warman, J. M. *Chem. Phys.* **73**, 35–53 (1982).
- Paddon-Row, M. N., Cotsaris, E. & Patney, A. K. *Tetrahedron* (in the press).
- Warman, J. M. & de Haas, M. P. *Interuniversity Reactor Institute Rep.* No. 134-85-07 (1985).
- Liptay, W. *Excited States* Vol. 1, 196–205 (Academic, New York, 1974).
- Holten, D., Windsor, M. W., Parson, W. W. & Thornber, J. P. *Biochim. biophys. Acta* **501**, 112–126 (1978).
- Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *J. molec. Biol.* **180**, 385–398 (1984).

Orientation-specific cortical responses develop in early infancy

O. J. Braddick, J. Wattam-Bell & J. Atkinson

Visual Development Unit, Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

Neurons in the visual cortex of higher mammals differ from those elsewhere in the visual pathway in that the majority respond selectively to particular edge or bar orientations in the stimulus^{1,2}. We have developed a visually evoked potential (VEP) technique which isolates the response of orientation-selective mechanisms from that of cortical or sub-cortical neurones which lack orientation selectivity. We are unable to find such orientation-selective responses in newborn human infants within the sensitivity of our method, but repeated longitudinal testing of individual infants shows that measurable responses emerge around 6 weeks of age. This result is consistent with the idea that human cortical visual function is very immature at birth, but develops rapidly in the first two postnatal months.

An evoked potential stimulated by changing the orientation of a fixed grating pattern would not demonstrate the existence of orientation-specific neurones, because the orientation change would introduce widespread local contrast changes in the display and could therefore activate neurones with non-oriented receptive fields. Our method isolates an orientation-specific response by embedding the orientation change in a dynamic sequence (see Fig. 1). A square-wave grating from a special-purpose digital signal generator is displayed on a video monitor. On every new video frame (25 Hz) the grating undergoes a random displacement (a phase shift in the range $\pm 5^\circ$ to $\pm 180^\circ$). On every third video frame (that is, 8.3 times per s) the orientation of the grating is changed. For a neurone with a circularly symmetrical receptive field, the change in stimulation when the orientation changes will be indistinguishable from the changes occurring with each random displacement. Thus, while such non-orientation-selective neurones may contribute a 25-Hz component to the VEP, any component at 8.3 Hz can arise only from neurones whose response depends specifically on orientation. (A small 25-Hz component is in fact visible in the response of adult subjects, but not in the infants we have tested.)

In testing infants, we have used alternation between orientations $\pm 57^\circ$ from vertical. Oblique orientations were chosen so that the common astigmatism of infancy^{3,4} (which is usually in the horizontal and vertical meridians) could not introduce effective contrast differences between gratings of the two orientations. The spatial frequency of the stimulus was 0.25 cycles deg^{-1} at the infants' viewing distance of 40 cm. Bipolar signals

from Ag/AgCl scalp electrodes atinion and vertex (with an indifferent electrode on the forehead) were fed via an isolated pre-amplifier and Biodata amplifier (overall gain 20,000; pass-band 0.2–30 Hz) to a microcomputer which acted as signal averager. Our methods have been described more fully elsewhere⁵. The computer cumulated the signal in blocks of 25 sweeps (each sweep including four stimulus reversals). Analysis of the VEP was normally based on a set of 10 such blocks.

It was necessary to be sure that any failure to record an orientation-specific VEP was not simply because we could not measure any response to gratings at the spatial and temporal frequencies used. We therefore also made recordings in a comparable pattern-appearance condition. In this, only one grating orientation was displayed, undergoing the same random displacements, and alternating at 4.16 Hz with a blank field of equal space-average luminance. If this stimulus can produce a measurable VEP in an individual infant, the spatio-temporal parameters of the grating cannot be held responsible for any absence of VEP to orientation reversal.

To test for the presence of a reliable VEP, we analysed the amplitude and phase of components at 4.16 and 8.33 Hz in each 25-sweep block of averaged signals. The consistency of the phase across blocks was tested using a version of the Rayleigh statistic⁶ which weights each measurement according to its amplitude^{7,8}. This statistic tests the data, at the $P < 0.05$ level, against the null hypothesis that no signal is present.

Having verified that adults gave clear VEP responses to the orientation-reversal stimulus, we tested a group of healthy newborn infants aged 0–9 days at Cambridge Maternity Hospital. Figure 2a shows a typical set of results. We only considered the orientation-reversal results for infants (such as that illustrated) with whom we could successfully record a statistically significant VEP to the pattern-appearance version of our display. In 20 such infants (17 born within 1 week of term; 3 who were 2–3 weeks pre-term), we did not see any instance of a statistically significant response to the orientation-reversal stimulus. To provide a more sensitive test for the possibility of a weak response to orientation reversal, we averaged signals for at least twice as long (500 sweeps) in three infants out of this group, and in two others who did not subsequently remain in an alert state long enough to complete pattern-appearance recording. In none of these did we see any indication of a reliable response to orientation reversal, despite the greater sensitivity of a prolonged averaging period. However, the orientation-reversal response seen at later ages is always markedly smaller than the pattern-appearance response (see Fig. 2b, c), so the possibility must remain that the newborn's immature visual system generates an orientation-reversal response which is too weak for the sensitivity of our methods.

As any orientation response would be expected to be at the second harmonic (8.33 Hz), while the pattern-appearance

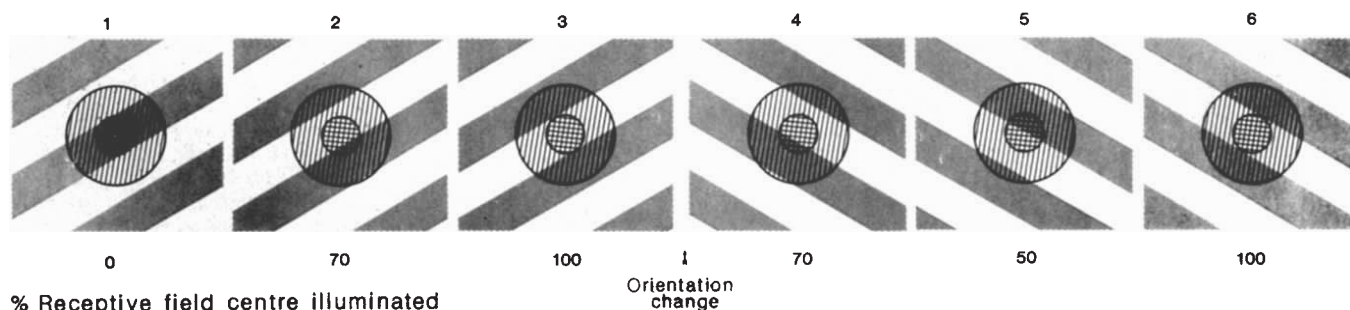


Fig. 1 A sample from the stimulus sequence used to elicit orientation-specific VEPs. Panels 1–6 show successive video frames superimposed on a hypothetical circularly symmetrical receptive field. With each new frame the grating undergoes a random displacement, and between frames 3 and 4 its orientation is reversed. The numbers below each panel show that, for the circular receptive field, the change in the percentage of the centre area illuminated when the orientation is reversed cannot be distinguished from the changes due to the phase shifts with each new frame. A VEP synchronized with the reversals is therefore evidence for orientation selectivity.

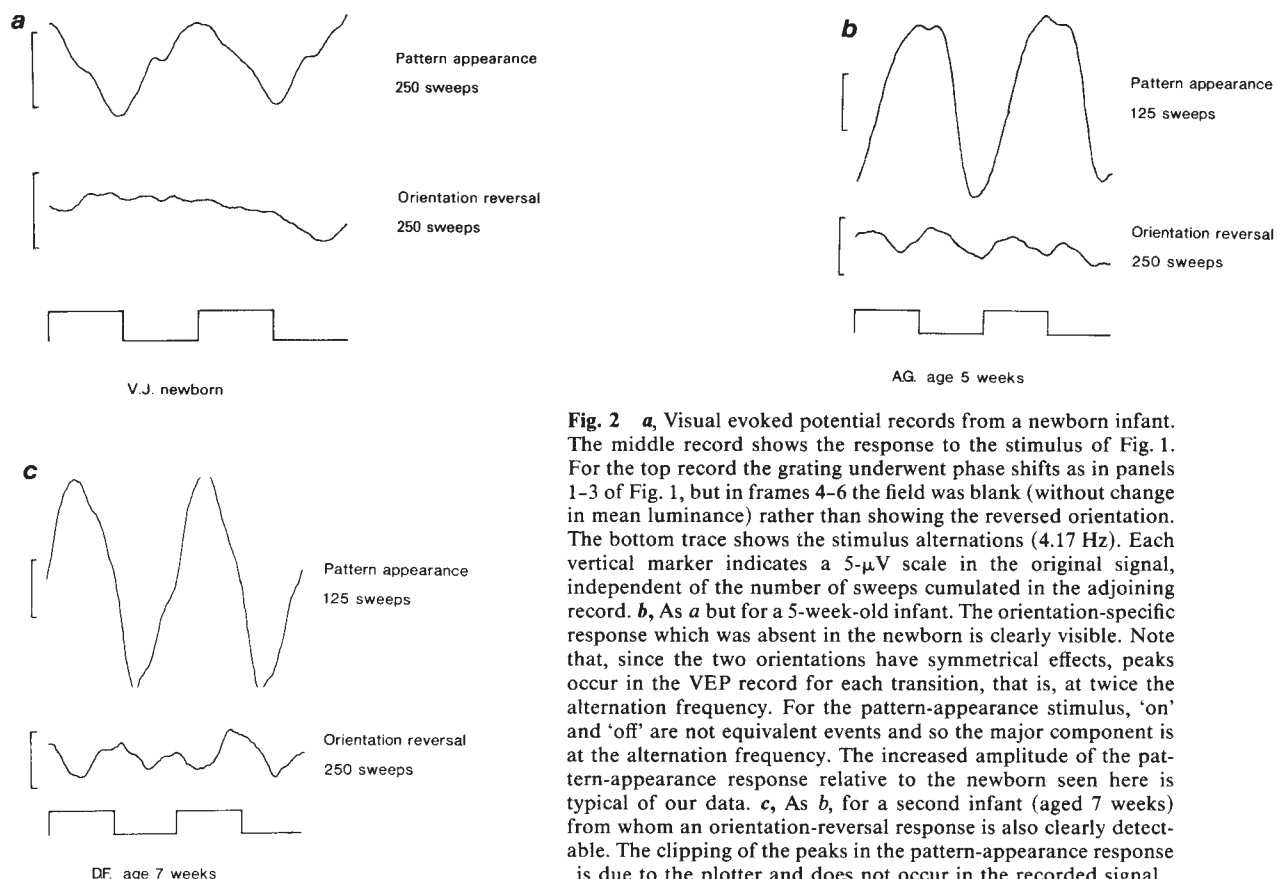


Fig. 2 *a*, Visual evoked potential records from a newborn infant. The middle record shows the response to the stimulus of Fig. 1. For the top record the grating underwent phase shifts as in panels 1–3 of Fig. 1, but in frames 4–6 the field was blank (without change in mean luminance) rather than showing the reversed orientation. The bottom trace shows the stimulus alternations (4.17 Hz). Each vertical marker indicates a 5- μ V scale in the original signal, independent of the number of sweeps cumulated in the adjoining record. *b*, As *a* but for a 5-week-old infant. The orientation-specific response which was absent in the newborn is clearly visible. Note that, since the two orientations have symmetrical effects, peaks occur in the VEP record for each transition, that is, at twice the alternation frequency. For the pattern-appearance stimulus, 'on' and 'off' are not equivalent events and so the major component is at the alternation frequency. The increased amplitude of the pattern-appearance response relative to the newborn seen here is typical of our data. *c*, As *b*, for a second infant (aged 7 weeks) from whom an orientation-reversal response is also clearly detectable. The clipping of the peaks in the pattern-appearance response is due to the plotter and does not occur in the recorded signal.

response is dominated by the fundamental frequency (4.16 Hz), we must consider whether the absence of the former in newborns might be to a limitation of the frequencies at which the immature brain can respond. Two lines of evidence argue against this. First, in an earlier study⁹ we have recorded conventional pattern-reversal responses from newborns at higher frequencies than 8 Hz. Second, three of our newborns showed a significant second-harmonic component in their pattern-appearance VEP (that is, they were demonstrably capable of 8.33-Hz responses), but none of these gave significant orientation responses at 8.33 Hz.

Overall, then, we conclude that this method provides no positive evidence of orientation selectivity in the visual cortex of the newborn infant.

In contrast, Fig. 2*b* and *c* show responses recorded from a 5-week-old and a 7-week-old infant; both pattern-appearance and orientation-reversal responses are clearly visible and statistically significant. To identify the age at which orientation-reversal VEPs could first be recorded, we studied a group of infants whom we first tested as close to 4 weeks of age as could be arranged, and then at approximately 2-weekly intervals. As with the newborns, the presence of a VEP for pattern appearance of the dynamic stimulus was verified with each infant included in the study. Figure 3 summarizes the results with this group. The median age at which an orientation-specific VEP response was first recorded was 6 weeks. (Given the intervals between the last negative and the earliest positive result, this may be a slight overestimate of the true median age of onset.)

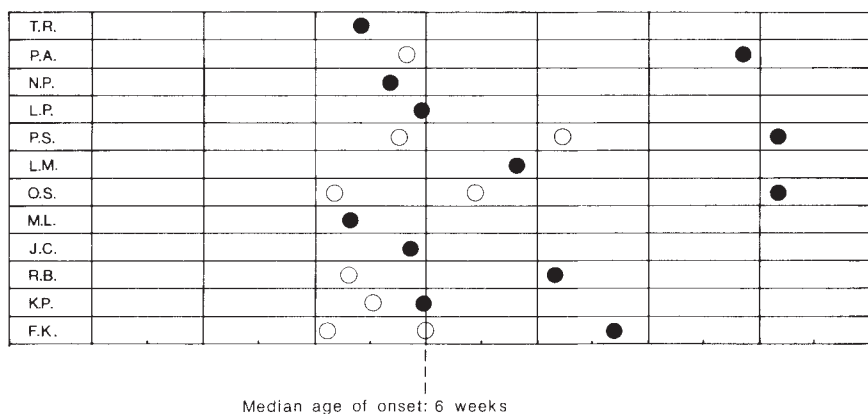


Fig. 3 Each row shows results of repeated testing of an individual infant with the orientation-reversal stimulus. An open circle indicates absence of any significant response at that age, while a solid circle indicates presence of a significant response. The single infant whose results are shown in the lower panel was born 3 weeks pre-term, with the scale referring to post-term rather than postnatal age.

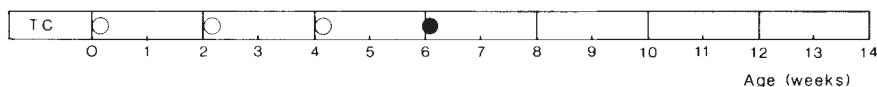


Figure 3 includes the results from one infant born significantly pre-term. While we do not wish to place much weight on results from a single individual, the finding that this child's first orientation-specific VEP response was found at 6 weeks post-term rather than 6 weeks postnatal is consistent with other findings that early visual development is not accelerated by visual experience before the due date of birth^{10,11}.

The first 2 months of human life show many striking qualitative developments in visual behaviour¹². Our results show the onset during this period of a response that is specifically associated with a key aspect of mature cortical function. They therefore support the view that vision before about 6 weeks of age may depend largely on subcortical pathways¹³, or at least, that any cortical processing before this age lacks important features of adult vision. Different aspects of cortical function do not necessarily mature together; using closely analogous techniques, we have shown that binocular input to cortical neurones is not found until about 13 weeks on average⁵.

The behavioural tests^{14,15} most closely relevant to the present VEP test have shown the ability to discriminate obliques at 6 weeks, but did not test younger infants. However, there are reports of behavioural discriminations involving orientation^{16,17} before 6 weeks. We do not know whether these discriminations are mediated by orientation-selective neurones that for some reason the VEP method does not detect, or whether they depend on other means of discrimination (such as directional biases in the infants' eye movements).

Single-unit recordings from the striate cortex of newborn monkeys show a significant number of cells possessing orientation selectivity¹⁸. There are several possible explanations of this apparent discrepancy with our findings in infants. First, the VEP method may only pick up a specific subpopulation of orientation-selective cells. For instance, if the earliest orientation-selective cells to develop were simple cells showing strictly linear spatial summation¹⁹, the random phase jumps of our stimulus would produce both positive and negative response changes in such cells and the response to orientation changes would not stand out from this range (however, a response which is truly linear in this respect is probably unrealistic in the cortex). Second, the relatively high rate of stimulation used in this study might yield different results from the lower frequencies typically used in primate single-unit recording. Third, it is possible that in early infancy, neurones specific to horizontals and verticals may predominate over those responding to the obliques we tested, as has been reported in other species^{20,21}. Finally, there are several lines of evidence that, compared with other primates, human infants are neurologically relatively immature at birth²²; cortical orientation selectivity may be a further aspect of this. Further experimental research, using both VEP and complementary behavioural techniques, may help us to specify more exactly in what, if any, respects the human visual cortex is functional at birth.

We thank the parents of our infant subjects for making these studies possible, Dr N. C. R. Robertson and the staff of Cambridge Maternity Hospital for their cooperation, and Shirley Anker, Carol Evans and Rose Anne Hall for assistance. This work is supported by an MRC programme grant.

Received 23 October 1985; accepted 26 February 1986.

- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106-154 (1962).
- Hubel, D. H. & Wiesel, T. N. *Proc. R. Soc. B* **198**, 1-59 (1977).
- Howland, H. C., Atkinson, J., Braddick, O. J. & French, J. *Science* **202**, 331-333 (1978).
- Mohindra, I., Held, R., Gwiazda, J. & Brill, S. *Science* **202**, 329-331 (1978).
- Braddick, O. J., Wattam-Bell, J., Day, J. & Atkinson, J. *Hum. Neurobiol.* **2**, 65-69 (1983).
- Greenwood, A. & Durand, D. *Ann. math. Statist.* **26**, 233-246 (1955).
- Moore, B. R. *Biometrika* **67**, 175-180 (1980).
- Wattam-Bell, J. *Perception* **14**, A33 (1985).
- Atkinson, J., Braddick, O. & French, J. *Invest. Ophthalmol. vis. Sci.* **18**, 210-213 (1979).
- Dobson, V., Mayer, D. L. & Clifton, P. L. *Invest. Ophthalmol. vis. Sci.* **19**, 1498-1505 (1980).
- van Hof-van Duin, J. & Mohn, G. in *Continuity of Neural Functions from Prenatal to Postnatal Life* (ed. Prechtl, H. F. R.) 93-114 (Blackwell, Oxford, 1984).
- Atkinson, J. *Hum. Neurobiol.* **3**, 61-74 (1984).
- Bronson, G. W. *Child Dev.* **45**, 873-890 (1974).
- Maurer, D. & Martello, M. *Vision Res.* **20**, 201-204 (1980).

- Cohen, L. B. & Younger, B. A. *Infant Behav. Dev.* **7**, 37-47 (1984).
- Fantz, R. L. & Miranda, S. B. *Child Dev.* **46**, 224-228 (1975).
- Slater, A. M. & Sykes, M. *Child Dev.* **48**, 545-554 (1977).
- Wiesel, T. N. & Hubel, D. H. *J. comp. Neurol.* **158**, 307-318 (1974).
- Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. *J. Physiol., Lond.* **283**, 53-77 (1978).
- Fregnac, C. Y. & Imbert, M. *J. Physiol., Lond.* **278**, 27-44 (1978).
- Leventhal, A. G. & Hirsch, H. V. B. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1272-1276 (1977).
- Prechtl, H. F. R. (ed.) *Continuity of Neural Functions from Prenatal to Postnatal Life* (Blackwell, Oxford, 1984).

Visual discrimination of target displacement remains after damage to the striate cortex in humans

Isobel M. Blythe, Jane M. Bromley, C. Kennard* & K. H. Ruddock†

Departments of Physics (Biophysics) and Pure and Applied Biology, Imperial College, London SW7 2BZ, UK

* Department of Neurology, London Hospital, Whitechapel, London E1 1BB, UK

Damage to the striate cortex usually causes blindness in those regions of the visual field which map to the area of neural damage^{1,2}. Nonetheless, there are reports that some patients with such damage can localize and perform certain visual discriminations between light stimuli presented within the 'blind' area of the visual field³⁻⁶. Experiments on animals with different brain areas ablated suggest that visual function is served by two principal projection pathways from the retina^{7,8}. That to the striate cortex is primarily responsible for fine discrimination between stimulus parameters such as colour and spatial pattern, whereas that to the superior colliculus in the midbrain is responsible for visual localization of stimuli. The residual visual functions in patients with cortical damage are usually attributed to the non-striate retinal projection to the superior colliculus. We now present measurements of spatial discrimination in two observers with large visual field defects (scotomata) caused by damage to the striate cortical region. Both exhibit a near normal ability to discriminate displacements of targets when two lights are flashed sequentially in their defective visual field, but they are unable to discriminate spatial pattern or size. We argue that these results are consistent with the 'two visual systems' interpretation of ablation studies on non-human species^{7,8}.

The representation of the visual field on the human striate cortex was first established from the study of patients who developed permanent blindness in localized areas following gunshot wounds to the striate cortex¹. However, visual discriminations in response to light stimulation within 'blind' areas of the visual field have been reported^{3-6,9} and two qualitatively different responses have been described. In one, visual discrimination can be shown only as forced-choice judgements between alternatives and the subject has no conscious visual sensation; these responses are thus termed 'blindsight'⁵. In the other, subjects report a sensation, usually described as a 'dark shadow', in response to bright, transient lights presented within visual field areas classed as 'blind' on the basis of conventional perimetry using steady or low-illumination transient stimuli⁹. The two subjects of the present study show the latter response, and the data presented relate to conscious detection of stimulus lights.

G.Y., a 28-year-old male, is hemianopic, suffering homonymous loss of vision in the right hemifield with sparing of about 3° around the fovea. R.C., a 27-year-old male, has a field loss covering the upper right quadrant, homonymous for the two eyes. The first subject received brain injuries in a motor accident, while the second sustained a cortical infarct following an attack of migraine. Degeneration of the left striate cortex, total for G.Y. and partial for R.C., has been established by computer

† To whom correspondence should be addressed.

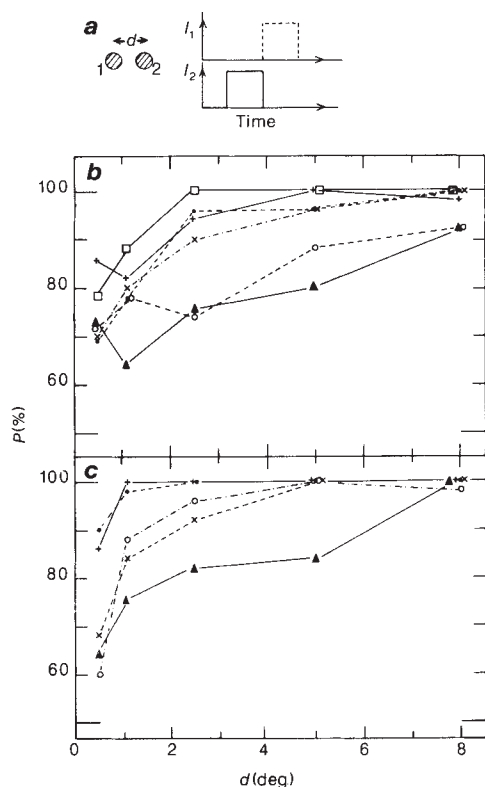


Fig. 1 *a*, The stimulus configuration used to elicit apparent displacement. The two circular targets were displaced centre to centre by distance d which was the variable parameter of the measurements, and were presented successively for 800 ms each. The value of d and the order sequence of the two spots were varied randomly between presentations. Each stimulus configuration was repeated 10 times. I_1 and I_2 denote the (equal) illumination levels of the two targets. *b*, Data of subject G.Y.; probability for correct discrimination of the target, $P(\%)$, is plotted against d , the separation of the two spots, expressed in degrees. The spots were of 1° diameter and the illumination level was 3.2 log Troland. Measurements were made in the 'blind' hemifield, at 5° (+), 10° (●), 15° (x), 20° (○) and 30° (▲) off-axis along the horizontal meridian, and at 5° off-axis in the normal hemifield (□). *c*, As for *b*, but for the normal subject I.B.; symbols are as for *b*.

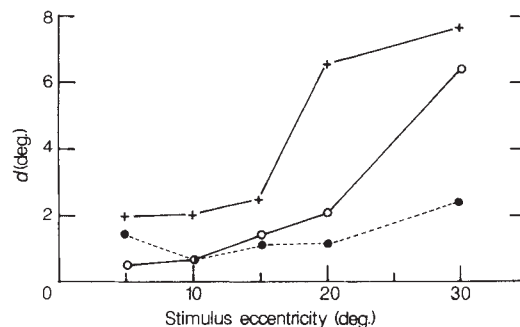


Fig. 2 *a*, Values of d required for 90% probability of correct discrimination of the direction of target displacement from right to left, or vice versa. Data for subject G.Y., blind hemifield (+) and I.B. (○), plotted against stimulus eccentricity, are taken from Fig. 1. Data for G.Y.'s normal hemifield (●) were obtained by similar methods.

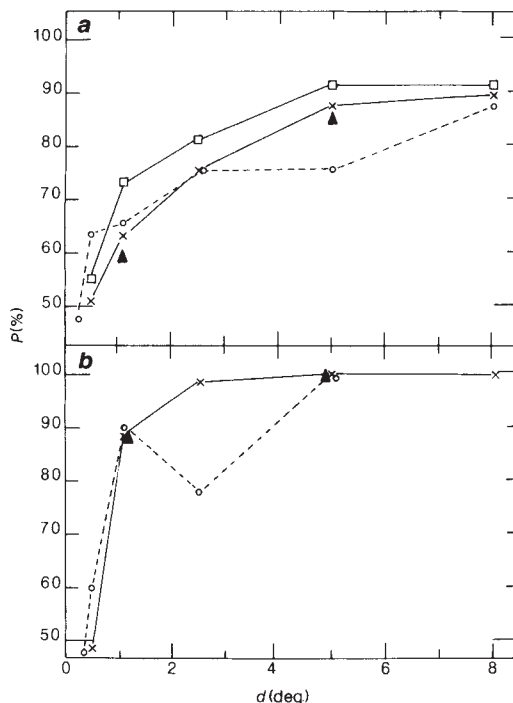


Fig. 3 *a*, As Fig. 1, but measurements were made at a single retinal location, 10° off-axis relative to both the vertical and horizontal meridians, placing the target in the upper right quadrant, that is, within R.C.'s 'blind' field area. The data for R.C. refer to circular targets (Fig. 1a) of 0.2° (▲), 0.5° (○) and 1.0° (x) diameter. Comparison data are also given for an equivalent location in the normal left upper quadrant for targets of diameter 1° (□). *b*, As *a*, but for the upper right quadrant of the normal subject I.B. Data refer to spots of diameter 0.2° (▲), 0.5° (○) and 1.0° (x).

tomography X-ray scans of the brain. Objects lying within these scotomata are detected only if displaced rapidly or presented transiently, and in the former case, the subjects experience a sensation of movement. Both G.Y. and R.C. perform accurate eye movements in order to fixate foveally on a flashed stimulus presented within the 'blind' hemifield, but neither can identify the colour, size or shape of the stimulus. For example, movement of the experimenter's hand cannot be differentiated from movement of a black, square card of much greater area. We have studied systematically the responses of the two subjects elicited by transient, white-light stimuli, provided by a maxwellian view optical system. All measurements were made monoptically, with the right eye, and the observer's head was kept steady relative to the instrumental exit pupil by a dental clamp.

Spatial discrimination was measured for the displacement of two spots, with centre-to-centre separation d in a direction parallel to the horizontal meridian, presented sequentially for 800 ms each (Fig. 1a). Such stimuli normally elicit a strong sensation of movement, directed from the first to the second flash, and this was reported by the two subjects for stimulus presentation within either the normal or 'blind' areas of the visual field. The probability P of correctly identifying spot displacement to the right or to the left was measured as a function of spot separation d . Data for subject G.Y., measured at several retinal locations, show that P rises rapidly as d increases, and

that the value of d for a given value of P increases as stimulus eccentricity increases (Fig. 1b). Values for G.Y.'s 'blind' hemifield are somewhat larger than those for his normal hemifield and larger than those measured for a normal subject I.B. (Fig. 1c), the differences being more significant at higher eccentricities (Fig. 2).

Data for R.C. are given for a single stimulus location, but for different spot sizes, and his responses are essentially independent of spot size (Fig. 3a). These data are again similar to those obtained for the normal field (Fig. 3a) and for the normal subject I.B. (Fig. 3b), although, as with G.Y., R.C.'s 'blind' field yields lower discrimination for displacement than does I.B.'s normal field.

For the 'blind' field, neither subject was able to resolve the targets used in the determination of displacement thresholds.

For example, G.Y. was presented with either a single circular target (diameter 1.1°) or with two such targets, separated spatially by 7° (as in Fig. 1a) but coincident in time, both targets being centred 13.5° off-axis along the horizontal meridian. The illumination of the single target was double that of each component, and so the two stimuli could not be distinguished on the basis of total flux. G.Y. was instructed as to the nature of the two stimuli, and in a random sequence of 100 presentations, he correctly identified the single target on 23 of 50 presentations and the pair on 26 of 50 presentations, that is, at chance level. He achieved 100% correct discrimination for presentation of the targets in the corresponding region of his normal hemifield.

The subjects were also unable to identify geometrical patterns formed by sequentially flashed spots, for example, by three spots (1.1° diameter), flashed for 800 ms and separated centre to centre by 5° (that is, a triangle with a spot at each corner), presented in the 'blind' field. The subjects reported the sensation of movement along a curved line, but the triangle was not identified; thus, discrimination of displacement cannot be exploited to provide pattern discrimination. Further, neither subject could discriminate at greater than chance level targets of different shapes or sizes presented within their visual field scotoma, as long as different targets were matched in total flux. For example, at no point in their 'blind' fields could they discriminate between circles of diameter 4° and equilateral squares and triangles constructed from lines of length 4° , the different targets being presented in random sequence for 300 ms at 2-s intervals. They were also unable to distinguish between a square-waveform grating of 1 cycle deg^{-1} periodicity and a uniform field of illumination equal to the average illumination of the grating. It is difficult to compare such complete lack of spatial vision with normal function, but as an example, gratings of $10 \text{ cycles deg}^{-1}$ located 7° off-axis can be resolved in normal vision, whereas

under the same conditions, neither subject was able to resolve a 1 cycle deg^{-1} grating located in the 'blind' field.

Computer tomography X-ray scans show that in both subjects brain damage is restricted to the striate cortex, thus lack of pattern recognition or discrimination is attributed to disturbance of this area, which receives direct input from the retino-geniculate pathway. It has been argued that the characteristics of residual vision in one of these two subjects are similar to those of electrophysiological responses recorded from the superior colliculus⁹, which itself projects via the pulvinar to the prestriate cortex. Subsequently, however, a direct, albeit sparse, projection from the primate lateral geniculate nucleus to the prestriate cortex has been demonstrated^{10,11}. Although we cannot identify the non-cortical pathway which mediates the residual vision described here, our results demonstrate a complete functional dissociation between visual localization of transients and pattern discrimination, as would be predicted by the 'two systems theory' of visual organization^{7,8}.

We thank G.Y. and R.C. for voluntarily carrying out an extended series of experiments, and acknowledge the award of a research grant by the Jules Thorn Charitable Trust and that of an MRC research training award to J.M.B.

Received 13 November 1985; accepted 28 February 1986.

1. Holmes, G. *Br. J. Ophthalmol.* **2**, 353-384 (1918).
2. Teuber, H. L., Battersby, W. S. & Bender, M. B. *Visual Field Defects after Penetrating Missile Wounds of the Brain* (Harvard University Press, 1960).
3. Riddoch, G. *Brain* **40**, 15-57 (1917).
4. Poppel, E., Held, R. & Frost, D. *Nature* **243**, 295-296 (1973).
5. Weiskrantz, L., Warrington, E. K., Sanders, D. M. & Marshall, J. *Brain* **97**, 709-728 (1974).
6. Perenin, M. T. & Jeannerod, M. *Neuropsychologia* **16**, 1-13 (1978).
7. Sprague, J. M. *Science* **153**, 1544-1547 (1966).
8. Schneider, G. *Science* **163**, 895-902 (1969).
9. Barbur, J. L., Ruddock, K. H. & Waterfield, V. A. *Brain* **103**, 905-928 (1980).
10. Fries, W. *Proc. R. Soc. B213*, 73-80 (1981).
11. Yukie, M. & Iwai, E. *J. comp. Neurol.* **201**, 81-97 (1981).

Analysis of genetic mosaics shows that the extra epidermal cell divisions in *Knotted* mutant maize plants are induced by adjacent mesophyll cells

Sarah Hake & Michael Freeling

Department of Genetics, University of California, Berkeley, California 94720, USA

The technique of 'fate-mapping' or 'mosaic mapping' takes advantage of genetically marked progenitor cells that confer a phenotypic difference in order to determine the cells' contribution to the final organism. Pioneered by the gynandromorph studies of Sturtevant¹, this technique has been used extensively to understand *Drosophila* development². For certain plant species, naturally occurring chloroplast mosaics have been used to follow cell lineages³. In other plant species, mosaics can be generated by using X-ray-induced somatic mutations^{4,5}. Here we have used X rays to create genetic sectors in maize that are lacking a mutation which alters leaf development—*Knotted* or *Kn*. From the sectors observed, we determined that the genotype of the mesophyll (internal cell layers) and not the genotype of the epidermis (outer layer) was critical for *Kn* expression, and that the mesophyll induced the epidermis to divide.

Knotted (map position 127 on chromosome 1L) is a locus in maize defined by five dominant mutants⁶ that specify unexpected divisions in cells near lateral veins of the leaf blade. The knots resulting from these divisions are hollow protrusions from the plane of the leaf and involve cell divisions in all layers of the leaf. The lateral veins are not distorted but extend up through the centre of the knots, implying that these *Kn*-induced cell

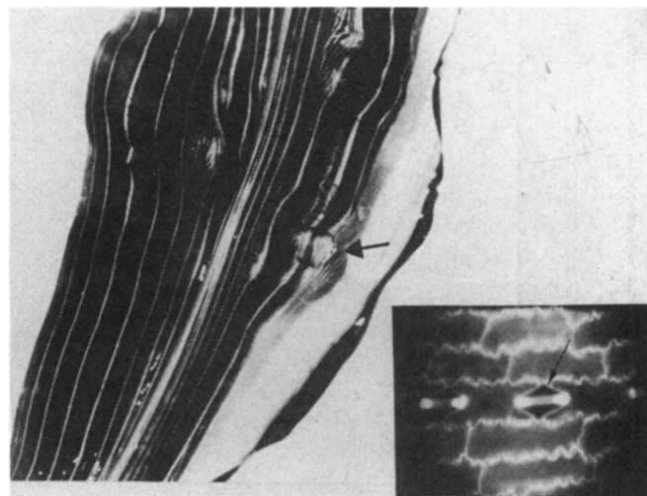


Fig. 1 Leaf sector of irradiated *Kn Lw/kn lw* plant 152. The white sector of this leaf has no knots; the light-green sector, in which some mesophyll layers remain *Kn*, is knotted. The arrow indicates a knot in the light-green sector. Inset, fluorescence photomicrograph of the epidermis of the white sector, showing that the guard cells contain chloroplasts (the guard cells are the only cells in the epidermis with chloroplasts). The arrow indicates a pair of guard cells. $\times 400$.

divisions occur early in leaf development before lateral veins differentiate⁷. We used X rays to create genetic sectors lacking the dominant *Kn* allele; these sectors were marked by uncovering a tightly linked locus that confers an absolutely white phenotype, the dysfunctional allele of the *lw* gene (position 128 on chromosome 1L). When seeds of the genotype *Kn Lw/kn lw* are irradiated, plants that lose the dominant alleles because

Table 1 Presence of chlorophyll used as a marker for the *Kn1* locus in epidermis and mesophyll layers of 68 *lw kn* sectors

Plant no.	Leaf no.	Sector size*	Upper epidermis	Presence of chlorophyll Upper mesophyll†	Lower mesophyll†	Lower epidermis	Presence of knots
1		1/20					
1		1/20	✓			✓	
3		1/36		✓	✓	✓	✓
4		1/10					
5		1/10					
5		1/10		✓	✓		✓
6‡		1/18					
7		1/19					
7		1/19		✓	✓		✓
8		1/35					
8	9	1/70					
8	12	1/10					
8	13	1/25					
10		1/10					
12				✓	✓		✓
12		1/20					
16		1/7	✓			✓	
16		1/37	✓	✓	✓	✓	✓
17§		1/25	✓			✓	✓
18		1/20	✓				
21		1/70					
23		1/44					
30	10	1/6	✓	✓		✓	
30	11	1/8	✓	✓		✓	
30	12	1/11	✓	✓	✓	✓	
31		1/10					
40		1/10					
41‡		1/26					
42		1/30	✓			✓	
42		1/13	✓	✓		✓	✓
43		1/30	✓	✓	✓	✓	✓
43			✓	✓	✓	✓	✓
46‡	11	1/12					
46‡	12	1/40					
47	13	1/22					
47	13	1/80		✓	✓		✓
47	14	1/50				✓	
47	14	1/100		✓	✓	✓	✓
48		1/32					
55		1/30	✓			✓	
57‡	12	1/20					
57	13	1/40				✓	
57	13	1/40	✓	✓	✓	✓	
61‡	12	1/5					
61	13	1/4					
61	13	1/22		✓	✓		
68	11	1/50					
68	11	1/50			✓		✓
68	13	1/55			✓		✓
68	13						
71	14	1/57					
71‡	15	1/30					
75		1/18	✓			✓	
75		1/18					
75		1/55		✓	✓		✓
75			✓		✓	✓	✓
102		1/4	✓				
102		1/8	✓	✓			
104		1/14					
106		1/36	✓			✓	
106		1/30	✓	✓	✓	✓	✓
108		1/2	✓			✓	
109‡		1/12					
114		1/40					
115		1/9					
119		1/25					
123		1/40	✓	✓	✓	✓	✓
125		1/50	✓	✓	✓	✓	✓
125		1/20	✓			✓	
152		1/9	✓			✓	
152		1/14	✓	✓		✓	✓

Seeds of the genotype *Kn1-N2 Lw/kn lw* were germinated for 2 days on 30 °C pads, irradiated with X rays (1,000 rad) at 50–80 rad min⁻¹, and planted in the field. Approximately 1,000 plants from 2 separate experiments were examined for *lw* sectors. Only sectors through very knotted leaves are included; thus the absence of knots in a sector is significant. All leaves examined were above the eighth node so the leaf examined had initiated after the time of irradiation. Only light-green sectors adjacent to white sectors were included. Leaf number is given only when there were consecutive sectors in very knotted leaves. Half of the plants had sectors in more than one leaf.

* The width of the leaf and of the sector were measured at their widest point. *Kn* distorts the width of the leaf, but sector sizes are in the same size range as in more typical genetic backgrounds^{4,5,8}. Sectors without width measurements were less than 1/40 of the total leaf width.

† Pale-green sectors contain chlorophyll in one or more mesophyll layers but have no chlorophyll in the central mesophyll layer.

‡ The presence of chlorophyll in the epidermis was determined from the position of the sector, not direct observation of chlorophyll fluorescence. White sectors at the margin do not have chlorophyll in their epidermal layers. Sectors not present at the margin do have chlorophyll in their epidermal layers.

§ Plant 17 is an apparent exception to the general finding that a *Kn* epidermis is incapable of inducing knots. A mitotic recombination event between *Kn* and *lw* could explain this exception, or perhaps *Kn* was active somewhere within the (non-green) vascular system.

|| Plant 30 had large yellow and white intermixed sectors, none of which showed knots.



Fig. 2 Leaf sector of irradiated *Kn Lw/kn lw* plant 108. The white sector of this leaf (Table 1) contained chloroplasts in the guard cells of both upper and lower epidermal layers. Along the margin of the white sector was a narrow green sector.

of chromosome-induced breakage, give rise to white non-*Kn* sectors in green knotted plants. Because 50 map units separate *Kn Lw* from the centromere, chromosome breaks are relatively frequent. At the time of irradiation the seedlings consist of approximately six embryonic leaves and an apical meristem that will elaborate the remaining 10–15 leaves and reproductive structures. The target cells in the apex usually give rise to sectors of 1/10 to 1/32 of each leaf's width; these sectors are often found in more than one leaf^{4,5}.

Two controls were set up to show that the *Kn* phenotype could be expressed in the genotype of the sectors. First, we found *Kn lw* recombinants expressing knots in pure white tissue, thus proving that the *Kn* phenotype does not require some component of normal green tissue⁶. Second, we showed that *Kn1-N2/-* plants, which are segmentally monosomic for the distal 80% of 1L, are still knotted although the knots are less severe than expected⁶.

The white sectors were often bordered by light-green or yellow sectors (Fig. 1); these border regions contain some mesophyll layers without and some with chlorophyll-containing cells. Although most leaf cell divisions are longitudinal, creating long files of cells, there are also transverse and periclinal divisions that increase the width and thickness of the leaf, respectively. The light-green sectors probably arose from transverse divisions of the mutant cell that elaborated the white sector. We took advantage of the distinct chlorophyll-containing cells of the epidermis, the guard cells, to determine the genotype of the epidermal cells in the white and pale-green sectors (using fluorescence microscopy; Fig. 1 inset). As noted previously⁸, the genotypes of the epidermis and mesophyll sometimes differ, suggesting that there is a hierarchy of apical layers within the meristem: there are cells in the apex that produce both epidermal and mesophyll cells, but also distinct layers that specify only epidermis or only mesophyll.

To determine which cell layer, if any, was responsible for the knots, 68 *lw* sectors throughout very knotted leaves were examined (Table 1). Our results show that, with one exception, the white sectors lacked knots even when the epidermis carried the *Kn* gene (Figs 1, 2): 14 white sectors had chloroplasts in the epidermis, yet lacked knots. The presence of knots in the exceptional white sector (see plant 17, Table 1) could be explained by a mitotic recombination event between *lw* and *Kn* (these are separated by 1.5 meiotic map units). When pale-green sectors (contiguous with white sectors) were knotted, the genotype of the epidermis was not critical. The leaf of plant 75 serves as an example: a white sector that was completely free

of knots was bordered by pale-green knotted sectors. Half of the white sector had chloroplasts in both epidermal layers, the other longitudinal half did not. One of the green knotted borders did not have chloroplasts in either epidermis, whereas the other did. We found plants having sectors with chlorophyll in either upper mesophyll (plant 42) or lower mesophyll (68 and 75) that also contained knots, illustrating that both layers are capable of inducing surrounding non-*Kn* cells to participate in the cell divisions that lead to knots. Thus the internal cells of the leaf, either upper or lower mesophyll layers or vascular components, are the site of *Kn* action.

Current notions of plant development emphasize the importance of a change in the plane of cell division during morphogenesis (see ref. 9 and references therein). Most cell divisions are anticlinal with the cell wall perpendicular to the surface. Periclinal divisions (the cell wall is parallel to the surface) are thought to anticipate the formation of new organs. Green and co-workers have documented the importance for organ formation of an alteration of the cell division plane followed by a change in cell polarity^{10,11}. A role for the epidermis in the formation of new organs has also been suggested^{12–15}. Gelinis *et al.*¹⁶ found that an unexpected periclinal division of the epidermis was the first detectable sign of knot formation. Further, the ligule, a fringe-like tissue resulting from periclinal divisions of the epidermis^{17,18}, can be induced *de novo* by *Kn*⁶. Thus it was suggested that the epidermis induced knot formation¹⁶. However, our results demonstrate that the *Kn* genotype of the epidermis has no impact on the knotted phenotype.

Interaction of genetically distinct epidermal and mesophyll layers in plants has been described only once to our knowledge¹⁹: Stewart and co-workers studied a chimaeric plant in which the epidermis was of one species and the mesophyll of another. The cellular characteristics examined, such as cell size, pigment and pH, maintained their independent phenotype, but organ characteristics such as flower shape, were the product of an interaction of the two genotypes.

Aneuploid studies have shown that *Kn* either encodes a new function in the leaf or vastly overproduces a normal leaf product (neomorphic mutation)⁶. We know that this product does not diffuse for long distances in a transverse direction because white sectors are conspicuously not knotted⁶.

The mosaic analyses presented here have shown that the epidermis is clearly induced to divide into knots and thus information must pass from internal *Kn* leaf cells to the non-*Kn* cells nearby. Whether or not the inducing substance is the product of the *Kn* gene, and whether it causes epidermal cells to divide out-of-plane specifically, remains to be determined.

We thank C. Fenczik for help in the field work and for figure preparation. This study was supported by NSF grant DCB-15631.

Received 5 November 1985; accepted 26 February 1986.

1. Sturtevant, A. H. *Z. wiss. Zool.* **135**, 323–356 (1929).
2. Hotta, Y. & Benzer, S. *Nature* **240**, 527–536 (1972).
3. Stewart, R. N. & Dermen, H. *Am. J. Bot.* **66**, 47–58 (1979).
4. Steffensen, D. M. *Am. J. Bot.* **55**, 354–369 (1968).
5. Coe, E. H. & Neuffer, M. G. in *The Clonal Basis of Development* (eds Subtelny, S. & Sussens, I.) (Academic, New York, 1978).
6. Freeling, M. & Hake, S. *Genetics* **111**, 617–634 (1985).
7. Bryan, A. A. & Sass, J. E. *J. Hered.* **32**, 343–346 (1941).
8. Poething, S. in *Contemporary Problems in Plant Anatomy* (eds White, R. A. & Dickinson, W. C.) (Academic, New York, 1984).
9. Steffensen, D. M. *Am. J. Bot.* **55**, 354–369 (1968).
10. Green, P. B. & Lang, J. M. *Planta* **151**, 413–426 (1981).
11. Selker, J. M. L. & Green, P. B. *Planta* **160**, 289–297 (1984).
12. Lyndon, R. F. *Ann. Bot.* **49**, 281–290 (1982).
13. Green, P. B. *Rev. Pl. Physiol.* **31**, 51–82 (1980).
14. Ledin, B. *Am. J. Bot.* **41**, 11–17 (1954).
15. Abbe, E. C., Phinney, B. O. & Baer, D. F. *Am. J. Bot.* **38**, 744–751 (1951).
16. Gelinis, D., Postlethwait, S. N. & Nelson, O. E. *Am. J. Bot.* **56**, 671–678 (1969).
17. Sharman, B. C. *Ann. Bot.* **6**, 245–282 (1942).
18. Hake, S., Bird, R. McK., Neuffer, M. G. & Freeling, M. in *Plant Genetics* (ed. Freeling, M.) (Liss, New York, 1985).
19. Stewart, R. N., Meyer, F. G. & Dermen, H. *Am. J. Bot.* **59**, 515–524 (1972).

A gradient of cytoplasmic free calcium in growing rhizoid cells of *Fucus serratus*

C. Brownlee & J. W. Wood

Marine Biological Association of the United Kingdom,
The Laboratory, Citadel Hill, Plymouth, Devon PL1 2PB, UK

It is becoming increasingly clear that cytoplasmic Ca^{2+} ($[\text{Ca}^{2+}]_{\text{cyt}}$) has an important role in the regulation of plant cell functions as well as in animal cells¹⁻³. However, there is an acute lack of measurements of the cytoplasmic Ca^{2+} concentration ($[\text{Ca}^{2+}]_{\text{cyt}}$) in plants. Direct measurements have so far been limited to *Chara* and *Nitella*, using aequorin⁴, and to *Haemanthus* endosperm cells, using Quin-2 (ref. 5). The latter study demonstrated a gradient of $[\text{Ca}^{2+}]_{\text{cyt}}$ in dividing cells. Here we evaluate the use of Ca^{2+} -selective microelectrodes and the fluorescent indicator Quin-2 to measure $[\text{Ca}^{2+}]_{\text{cyt}}$ in rhizoids of germinating *Fucus serratus* zygotes. This preparation is particularly attractive for such an investigation as the *Fucus* zygote is a well-studied developmental system and the relatively large polarized rhizoid cells are predominantly cytoplasmic with no large vacuoles. We demonstrate the presence of a longitudinal gradient of $[\text{Ca}^{2+}]_{\text{cyt}}$ in the rhizoid cell. This gradient appears to be maintained by preferential Ca^{2+} influx in the region of the growing tip. The significance of the gradient for cell polarity is discussed.

F. serratus gametes were liberated in filtered sea water⁶ and zygotes were germinated on coverslips under continuous unilateral light at 17 °C for 6–24 h. Figure 1 shows typical results obtained with Ca^{2+} electrodes in extreme-apical (a) and sub-apical (b) regions of the rhizoid. The membrane potential (V_m) and resting calcium potential (V_{Ca}) recovered after impalement to give steady values. Recovery time varied from a few seconds to several minutes and was probably dependent on the degree of damage caused by the Ca^{2+} electrode as impalement with a single sharp V_m electrode usually gave steady V_m values within seconds. No consistent relation was found between V_m and V_{Ca} recovery patterns. Recordings could often be made up to 30 min. A frequent problem with the Ca^{2+} electrodes was displacement of the sensor from the electrode tip on impalement, probably caused by cell turgor pressure; this could usually be overcome by using zygotes that had not yet divided and which probably had lower turgor, and by ensuring that the sensor was as viscous as possible.

Table 1 shows that $[\text{Ca}^{2+}]_{\text{cyt}}$ measured at the extreme tip was higher than that measured in the sub-tip region. V_m values in

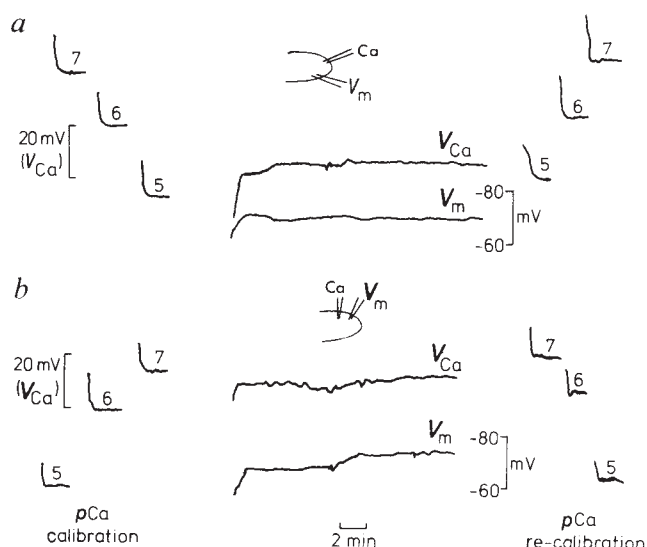


Fig. 1 Resting calcium potential (V_{Ca}) and membrane potential (V_m) in tip (a) and sub-tip (b) regions of the rhizoid of 1-day-old *F. serratus* zygotes in ASW (in mM): NaCl 500, KCl 10, CaCl_2 10, MgCl_2 10, NaHCO_3 2.5 pH 8.0. Calculation of $[\text{Ca}^{2+}]_{\text{cyt}}$ from these traces (see Table 1) yields values of 3.24 μM for tip and 0.18 μM for sub-tip regions.

Methods. Neutral carrier Ca^{2+} -selective microelectrodes were used^{21,22}. Fine electrodes were pulled from cleaned²² glass tubing (GC120TF, Clarke Electromedical and the tips were broken under a microscope to 0.5–1.0 μm diameter. These electrodes were silanized with tributylchlorosilane vapour²². A viscous column of Ca^{2+} sensor (Ca^{2+} cocktail (Fluka) + 16% polyvinyl chloride dissolved in tetrahydrofuran (THF)) was blown into the tip of the electrode via fine capillary. The electrode was left in a dessicator for 2 h during which excess THF evaporated, leaving a 100–200- μm column in the tip. Electrodes were back-filled with electrolyte²² and a fine hand-drawn capillary filled with electrolyte was pushed through the sensor column into the electrode tip. This short-circuited the bulk of the sensor and ensured good electrical contact between sensor and electrolyte and reduced the electrode response time. Calibration was in Ca^{2+} buffer solutions²². Available evidence²³ suggests that these buffer solutions match reasonably the major ions in *Fucus* cytoplasm. NaCl (20 mM) was added to the calibration solutions to match the probable cytoplasmic Na^+ concentration of *Fucus* zygotes²³. V_{Ca} was followed with a Burr-Brown 3431J amplifier. Electrodes were selected for sensitivity below pCa-7 and for minimal drift and hysteresis. V_m was recorded using electrodes filled with 3 M KCl. Ca^{2+} electrodes were inserted either in the extreme tip region (a) or ~20–30 μm behind the tip (b). V_m electrodes were inserted as close as possible to the Ca^{2+} electrodes. V_m was subtracted electronically from the total Ca^{2+} electrode potential to give V_{Ca} . Ca^{2+} electrodes were recalibrated after each impalement, results were discarded if electrode characteristics changed significantly on recalibration or if low V_m values were obtained.

Table 1 Values of $[\text{Ca}^{2+}]_{\text{cyt}}$ and V_m for rhizoid tip and sub-tip regions

	$[\text{Ca}^{2+}]_{\text{cyt}}$ (μM)	V_m (mV)
Tip	2.47 ± 0.8	-70 ± 11
Sub-tip	0.28 ± 0.06	-74 ± 9

Values for $[\text{Ca}^{2+}]_{\text{cyt}}$ and V_m represent means of 10 separate microelectrode experiments, $\pm$ s.e. In each experiment, steady recordings were followed for at least 15 min (Fig. 1). When stable potentials were obtained, pCa and V_m were sampled at 1-min intervals to give mean values for each experiment.

both regions ranged from -70 to -80 mV. These differences in $[\text{Ca}^{2+}]_{\text{cyt}}$ may represent the presence of a gradient of $[\text{Ca}^{2+}]_{\text{cyt}}$ or may be artefacts of the technique. It is possible that the Ca^{2+} electrode causes some membrane damage, which may be greater in the tip region where the membrane may be more susceptible to damage, resulting in an increase in $[\text{Ca}^{2+}]_{\text{cyt}}$. However, V_{Ca} and V_m quickly reached steady values in the experiments reported. In some instances in the tip region, a V_m of -80 mV was recorded together with a $[\text{Ca}^{2+}]_{\text{cyt}}$ of >1 μM . $[\text{Ca}^{2+}]_{\text{cyt}}$ was

much more variable in the tip region; this may be due to different degrees of penetration of the Ca^{2+} electrode. It was not possible to precisely locate the electrode tip in the cell. If $[\text{Ca}^{2+}]_{\text{cyt}}$ is higher in the tip region and decreases rapidly within a few micrometres, high $[\text{Ca}^{2+}]_{\text{cyt}}$ values would only be recorded when the electrode tip was just inside the plasmalemma. Measurements made with double-barrelled Ca^{2+} -selective microelectrodes have yielded $[\text{Ca}^{2+}]_{\text{cyt}}$ values of 2.61 ± 0.9 and 0.35 ± 0.15 μM (mean $\pm$ s.e.) for tip and sub-tip regions respectively ($n=4$). The $[\text{Ca}^{2+}]_{\text{cyt}}$ values in sub-tip regions are similar to those found in *Chara* cytoplasm⁴ and in several animal cells (for example, *Xenopus* and *Ambystoma* oocytes, measured with Ca^{2+} electrodes⁷). Micromolar resting $[\text{Ca}^{2+}]_{\text{cyt}}$ such as that found in tip regions has been measured in *Nitella* cytoplasm⁴.

The high $[\text{Ca}^{2+}]_{\text{cyt}}$ values in the tip region were further investigated in the experiments of Fig. 2. Removal of external Ca^{2+}

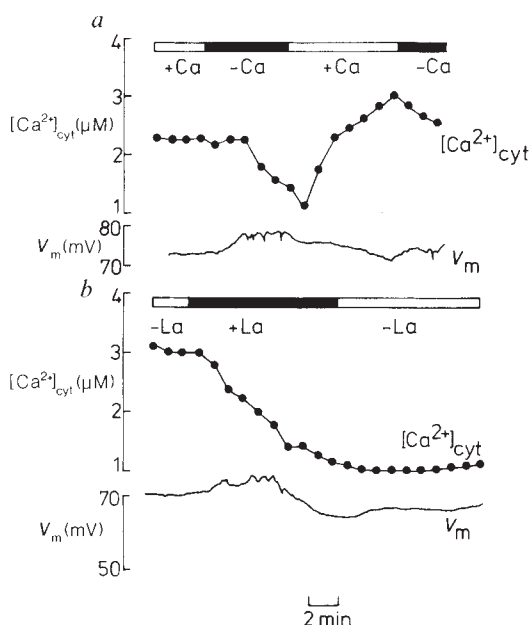


Fig. 2 *a*, Effect of removal of external Ca^{2+} on resting $[Ca^{2+}]_{cyt}$ and V_m in the rhizoid tip region of a 1-day-old *F. serratus* zygote. Cells in ASW were impaled with Ca^{2+} -selective and V_m microelectrodes. After steady values of V_{Ca} and V_m had been obtained, ASW was substituted for Ca^{2+} -free ASW (ASW with no $CaCl_2$, plus 100 mM $MgCl_2$). The effect of re-addition and removal of Ca^{2+} is also shown. *b*, Effect of 5 mM $LaCl_3$ on $[Ca^{2+}]_{cyt}$ and V_m . Cells were bathed in ASW without HCO_3^- . For both experiments, V_{Ca} values at 1-min intervals were converted to $[Ca^{2+}]_{cyt}$. Each experiment was performed three times.

(Ca^{2+}_{ext}) caused a steady reduction of $[Ca^{2+}]_{cyt}$ (Fig. 2a), and replacement of Ca^{2+}_{ext} caused a somewhat faster increase in $[Ca^{2+}]_{cyt}$. The effect was reversible. In this experiment, V_m hyperpolarized on removal of Ca^{2+}_{ext} ; this effect was also reversible. Similar, though slightly smaller, changes were observed with artificial sea water (ASW) containing 10 mM $MgCl_2$. Possible reasons for the relatively small changes in $[Ca^{2+}]_{cyt}$ in response to large changes in Ca^{2+}_{ext} include incomplete removal of cell-wall Ca^{2+} , a contribution to Ca^{2+}_{cyt} from intracellular stores, or the presence of a saturable carrier for Ca^{2+} in the plasmalemma. Blocking of putative Ca^{2+} channels with 5 mM La^{3+} gradually decreased $[Ca^{2+}]_{cyt}$ to submicromolar levels (Fig. 2b); this effect was only very slowly reversible. This result argues strongly against high $[Ca^{2+}]_{cyt}$ values in the tip region being caused by a damage artefact. In calibration solutions, La^{3+} closely mimics Ca^{2+} , producing a positive shift in V_{Ca} (data not shown). In the cell, La^{3+} caused a negative shift in V_{Ca} , which suggests that little or no La^{3+} was entering the cell. Any La^{3+} which did enter the cytoplasm would tend to cause an underestimate of the $[Ca^{2+}]_{cyt}$ decrease. After an initial hyperpolarization, La^{3+} depolarized V_m by 10–20 mV in most experiments. Taken together, these results strongly suggest that Ca^{2+} enters the rhizoid tip through Ca^{2+} channels down its electrochemical gradient and that the longitudinal $[Ca^{2+}]_{cyt}$ gradient within the rhizoid is maintained, at least in part, by this asymmetric net influx.

Incubation of 1-day-old zygotes with 20 μM Quin-2 acetoxymethyl ester (Quin-2 AM) for 2 h caused a shift in the emission spectrum to longer wavelengths, suggesting that some degree of uptake and hydrolysis of the ester to free Quin-2 had occurred⁸. However, the spectral shift was incomplete, showing a peak at 470–480 nm instead of the predicted 491 nm. Incubation for longer periods caused little further shift; this could be the result of either incomplete hydrolysis of the ester or incomplete removal of excess Quin-2 AM during washing. Small particles of undissolved Quin-2 AM were sometimes seen attached to

zygotes during fluorescence microscopy even after washing. Zygotes in this case were carefully selected for the absence of any attached Quin-2 AM particles. It is possible that sequestration of the ester into vacuoles of more mature cells may prevent its hydrolysis. The presence of the cell wall may reduce uptake of the ester and the highly pigmented state of the thallus cells will quench fluorescence. Because of these uncertainties it is not possible to obtain accurate per cent saturation values for Ca^{2+} Quin-2. Nevertheless, the rhizoid cells are relatively transparent and lend themselves to fluorescence microscopy⁹, as suggested in Table 2. Incubation with Quin-2 AM caused increased fluorescence in both tip and sub-tip regions which was further increased significantly ($P = 0.05$) by the ionophore A23187 only in sub-tip regions. Mn^{2+} reduced the signal almost to autofluorescence levels. From this experiment we conclude that Quin-2 fluorescence was saturated in the tip region, suggesting $[Ca^{2+}]_{cyt}$ values of $>1 \mu M$ ^{8,9}. In the sub-tip region, Quin-2 was not saturated before the addition of A23187, suggesting $[Ca^{2+}]_{cyt}$ values $<1 \mu M$ (K_d for Ca^{2+} Quin-2 = 115 mM; ref. 8). These results agree at least qualitatively with Ca^{2+} electrode results.

Table 2 Relative fluorescence at 491 nm for rhizoid tip and sub-tip regions of 2-day-old zygotes

	Autofluorescence	+Quin-2	+A23187	+ Mn^{2+}
Tip	1.38 $\pm$ 0.03	2.28 $\pm$ 0.16	2.04 $\pm$ 0.13	1.62 $\pm$ 0.14
Sub-tip	1.72 $\pm$ 0.25	2.81 $\pm$ 0.2	3.37 $\pm$ 0.13	1.9 $\pm$ 0.15

Fluorescence of selected areas of thizoid tips was recorded with a Leitz fluorescence microscope and a Newvicon low-light CCTV camera (Javelin Electronics) operated in the manual-gain mode. Light from a 209-W mercury vapour lamp was passed through a 340-nm excitation filter and a $\times 40$ objective. Fluorescence was monitored at 491 nm. The signal from the camera was displayed on a TV monitor and stored on video. After recording fluorescence, zygotes were incubated in 3 ml of 20 μM Quin-2 AM in ASW for 90 min, followed by three washes in ASW. Excitation was for <0.5 s per measurement to minimize dye bleaching. Fluorescence was measured after each treatment (150 nM A23187, 30 μM $MnCl_2$) which lasted for 30 min. Fluorescence due to A23187 alone was measured in controls; this was subtracted from all treatments involving A23187. Values shown are pooled means from three separate experiments. The signal from the video or camera could be enhanced with a custom-built contrast enhancer to improve visibility. The brightness of selected areas of the rhizoid ($15 \times 15 \mu m$) was measured using a custom-built video integrator. The signal was compared with a manually adjustable reference in a brightness comparator. Bright areas of the picture opened a gate to admit pulses to a counter. Additional circuitry generated an adjustable window to select areas of the picture. Ten frames were sampled per illumination to average out variations in lamp output. Changes in brightness of a control area of the screen were also monitored to adjust for background fluorescence and lamp variability.

Zygotes of *Fucus* and *Pelvetia* are known to pass a current through themselves. Positive current enters at the rhizoid tip and leaves from more mature regions^{10,11}. Similar current patterns have been observed in several other tip-growing systems¹². Ca^{2+} influx is a component of the inward current in the *Pelvetia* rhizoid tip¹³. This influx is probably largely balanced by Ca^{2+} efflux at the thallus end of the zygote to prevent saturation of endoplasmic reticulum and other intracellular Ca^{2+} stores. Such an asymmetric transcellular Ca^{2+} flux has been demonstrated previously¹³. A consequence of this Ca^{2+} influx is likely to be increased $[Ca^{2+}]_{cyt}$ in the tip region¹². Evidence for gradients of total or membrane-bound Ca^{2+} exists for several tip-growing cells^{14–16}.

The gradient of $[Ca^{2+}]_{cyt}$ reported here could have important physiological functions including the regulation of cytoskeletal processes¹⁷, exocytosis¹⁸ and ATPase activity¹⁹. Calmodulin has been detected in several tip-growing plant cells²⁰; Hausser *et al.* found that the distribution of calmodulin did not always agree with that of chlorotetracycline-visualized Ca^{2+} and they suggested that restricted availability of Ca^{2+}_{cyt} might interact with calmodulin. The results of the present study support this view.

We thank Professor P. F. Baker and Dr Joy A. Umbach for helpful advice and discussions. This research was supported by NERC.

Received 6 January; accepted 3 March 1986.

- Anderson, J. M. & Cormier, M. J. *Biochem. biophys. Res. Commun.* **84**, 595-602 (1978).
- Roux, S. J. & Slocum, R. D. in *Calcium and Cell Function* Vol. 3 (ed. Cheung, W. Y.) 409-453 (Academic, New York, 1982).
- Salimath, B. P. & Marme, D. *Planta* **158**, 560-568 (1983).
- Williamson, R. E. & Ashley, C. C. *Nature* **296**, 647-651 (1982).
- Keith, C. H., Ratan, R., Maxfield, F. R., Bajer, A. & Shelanski, M. L. *Nature* **316**, 848-850 (1985).
- Quatrano, R. S. in *Experimental Marine Biology* (ed. Mariscal, R.) 303-346 (Academic, New York, 1974).
- Moreau, M., Vilain, J. P. & Guerrier, P. *Dev. Biol.* **78**, 201-214 (1980).
- Tsien, R. Y., Pozzan, T. & Rink, T. J. *J. Cell Biol.* **94**, 325-334 (1982).
- Rogers, J. R. et al. *FEBS Lett.* **161**, 21-27 (1983).
- Nuccitelli, R. & Jaffe, L. F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4855-4859 (1974).
- Nuccitelli, R. & Jaffe, L. F. *J. Cell Biol.* **64**, 636-643 (1975).
- Nuccitelli, R. *Mod. Cell Biol.* **2**, 451-481 (1983).
- Robinson, K. R. & Jaffe, L. F. *Science* **187**, 70-72 (1975).
- Gilkey, J. C. & Jaffe, L. F. *Biophys. J.* **16**, 110a (1976).
- Jaffe, L. A., Weissenel, M. H. & Jaffe, L. F. *J. Cell Biol.* **67**, 488-492 (1975).
- Reiss, H.-D. & Herth, W. *Protoplasma* **115**, 153-159 (1983); *Protoplasma* **97**, 373-377 (1978); *Planta* **146**, 615-621 (1979).
- Williamson, R. E. *Pl. Cell Envir.* **7**, 431-440 (1984).
- Whitaker, M. J. & Baker, P. F. *Proc. R. Soc. B* **218**, 397-413 (1983).
- Dieter, P. *Pl. Cell Envir.* **7**, 371-380 (1984).
- Hausser, I., Herth, W. & H.-D. *Planta* **162**, 33-39 (1984).
- Lanter, F., Steiner, R. A., Ammann, D. R. & Simon, W. *Analy. chim. Acta* **135**, 51-59 (1982).
- Tsien, R. Y. & Rink, T. J. *Biochim. biophys. Acta* **599**, 623-638 (1980).
- Allen, R. D., Jacobsen, L., Joaquin, J. & Jaffe, L. F. *Dev. Biol.* **27**, 538-545 (1972).

Minor but not major histocompatibility antigens of thymus epithelium tolerize precursors of cytolytic T cells

Harald von Boehmer & Katrin Hafen

Basel Institute for Immunology, Grenzacherstrasse 487, 4005 Basel, Switzerland

Treatment of fetal thymuses with 2-deoxyguanosine depletes these organs of many haematopoietic cells¹, and if such thymuses are transplanted into allogeneic athymic nude mice, intrathymic development of cytolytic T-lymphocyte precursors (CTL-P) occurs, including those which are specific for class I major histocompatibility complex (MHC) antigens expressed by the thymus epithelium². Thus, T cells from BALB/c (*H-2^d*) nude mice transplanted with allogeneic C57BL/6 (*H-2^b*) thymic epithelium can be stimulated *in vitro* to produce CTL specific for *H-2^b* class I MHC antigens². We report here that thymocytes and lymph node T cells from such mice are responsive in mixed leukocyte reaction in the absence of exogenous growth factors, indicating that lack of tolerance is manifest at the level of CTL-P and proliferating T cells. We also show that T cells from such mice are tolerant to minor histocompatibility antigens of the thymus donor in the context of MHC antigens of the recipient. The results indicate that haematopoietic rather than epithelial cells tolerize CTL-P and that donor-type minor but not major histocompatibility antigens can be presented in tolerogenic form by haematopoietic cells expressing recipient-type MHC antigens.

Thymocytes from BALB/c *nu/nu* mice transplanted with 2-deoxyguanosine-treated thymuses from either B6 or B10.BR mice respond in mixed lymphocyte reaction to anti-Thy-1-treated B6 and B10.BR cells, respectively, but not to BALB/c stimulator cells (Table 1). They respond better to third-party stimulator cells, but responses to donor-type cells are always significant. Figures 1, 3 and 4 show that lymph node or spleen cells from the same mice give primary CTL responses when stimulated *in vitro* with X-ray irradiated B6 or B10.BR cells. These experiments indicate that in the grafted mice also, T cells required for proliferative responses to donor antigens were not tolerized; that is, donor type MHC antigens cannot be presented in tolerogenic form by either the thymus epithelium or recipient *H-2^d* haematopoietic cells. This contrasts with results obtained in experiments in which thymuses were cultured at low temperatures before transplantation into nude mice³, and suggests

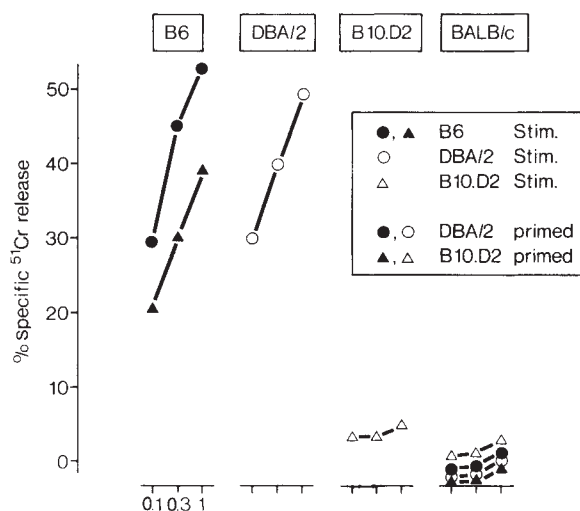


Fig. 1 Four-week-old BALB/c *nu/nu* mice were transplanted with six (three under each kidney capsule) 2-deoxyguanosine-treated¹ B6 thymuses. Eight weeks later, one grafted mouse was immunized with DBA/2 and another with B10.D2 spleen cells: 10^7 X-irradiated (2,200R) spleen cells were injected intraperitoneally (i.p.). Two weeks later, mesenteric lymph node cells were stimulated with B6, DBA/2 or B10.D2 stimulator cells as indicated. All cultures contained interleukin-2 (IL-2) except those containing B6 stimulators⁹. Cytolytic activity was tested after 5 days of culture according to methods described in detail previously⁹. The various target cells are indicated at the top of the figure.

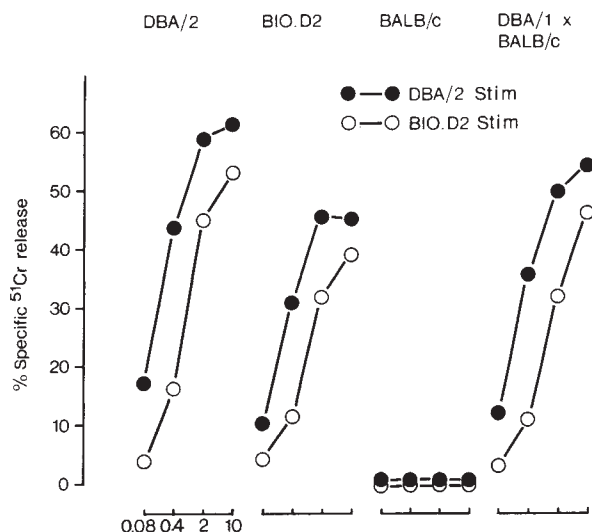


Fig. 2 A BALB/c mouse was injected i.p. with 10^7 X-irradiated (2,200R) spleen cells each of DBA/2 and B10.D2 mice. Two weeks later spleen cells were stimulated with either DBA/2 or B10.D2 spleen cells and assayed for cytolytic activity as described in Fig. 1 legend.

that a less rigorous depletion of haematopoietic cells may fail to remove the tolerizing cells. On the other hand, our results are compatible with *in vitro* experiments in which 2-deoxyguanosine-treated thymuses were repopulated by allogeneic thymic stem cells⁴.

Findings analogous to our results have been reported for the immunogenicity of MHC antigens: grafts lacking haematopoietic cells that differ in MHC antigens from the recipient often are not rejected, indicating that donor-type MHC antigens are not presented in immunogenic form by either the

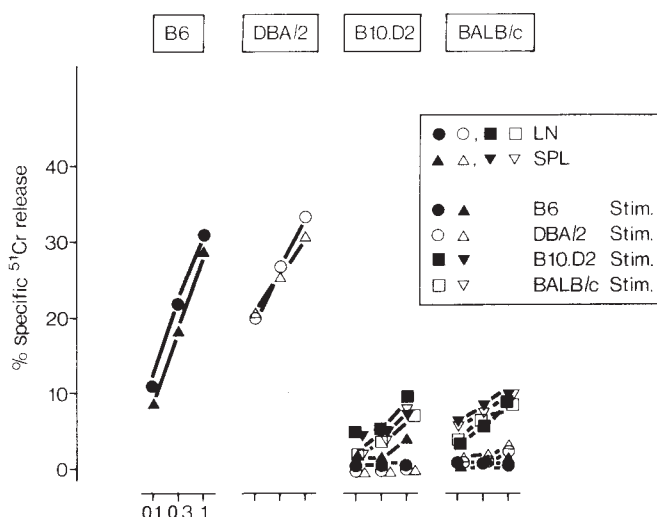


Fig. 3 A 4-week-old BALB/c *nu/nu* mouse was transplanted with six (three under each kidney capsule) 2-deoxyguanosine-treated¹ B6 thymuses. Eight weeks later the mouse was immunized with DBA/2 and B10.D2 spleen cells: 10^7 X-irradiated (2,200R) spleen cells were injected i.p. Two weeks later spleen cells (SPL) or mesenteric lymph node cells (LN) were stimulated with B6, DBA/2 or BALB/c stimulator cells as indicated. All cultures except those containing B6 stimulators contained IL-2⁹. Cytolytic activity was tested after 6 days of culture as described previously⁹.

grafted epithelial cells or the recipient haematopoietic cells^{5,6}. In contrast, minor histocompatibility antigens from grafts^{7,8} can be presented by haematopoietic cells⁹ of recipient type to prime minor histocompatibility antigen-specific CTL-P restricted by recipient-type MHC antigens. We analysed whether similar rules exist with regard to the tolerization of minor histocompatibility antigen-specific CTL-P in BALB/c *nu/nu* mice grafted with 2-deoxyguanosine-treated thymuses. If most B6 or B10.BR minor histocompatibility antigens were present on the transplanted thymus epithelium and presented in tolerogenic form in the context of H-2^d MHC antigens by recipient cells, we expect CTL-P from these mice would not be stimulated by B10.D2 cells (expressing donor-type minor and recipient-type major or medial¹⁰ histocompatibility antigens) but would respond to DBA/2 cells (expressing third-party minor and recipient-type major and medial histocompatibility antigens).

In a first experiment, one grafted BALB/c *nu/nu* mouse was primed with X-irradiated B10.D2 while another was primed

with X-irradiated DBA/2 spleen cells. Lymph node cells from both mice could be stimulated by B6 cells to generate B6-specific CTL. B10.D2-primed and *in vitro*-stimulated spleen cells did not produce CTL while DBA/2-primed and *in vitro*-stimulated spleen cells generated DBA/2-specific CTL (Fig. 1). The results indicate that the B6-grafted BALB/c nude mice were specifically tolerant to B6 minor histocompatibility antigens expressed on B10.D2 stimulators, that is, in the context of recipient MHC antigens, as BALB/c mice produce CTL after priming with DBA/2 and B10.D2 and stimulation with B10.D2 or DBA/2 cells (Fig. 2). It is especially important that CTL-P from DBA/2 plus B10.D2-primed BALB/c mice, induced by DBA/2 stimulators *in vitro*, lyse both DBA/2 and B10.D2 targets. CTL induced by B10.D2 stimulators *in vitro* also lyse both DBA/2 and B10.D2 targets. Neither of the CTL populations lyse DBA/1 targets, but they do lyse (DBA/1 $\times$ BALB/c)_{F1} targets, indicating that they are indeed H-2^d-restricted and specific for DBA minor histocompatibility antigens, some of which are also expressed on B10.D2 targets. A very different result is obtained when BALB/c *nu/nu* mice are grafted with 2-deoxyguanosine-treated thymuses from B6 or B10.BR mice, primed with DBA/2 plus B10.D2 cells and stimulated *in vitro* with DBA/2 or B10.D2 stimulators. Figures 3 and 4 show that spleen and lymph node cells from B6-grafted BALB/c *nu/nu* mice produce CTL after stimulation with DBA/2 but not B10.D2 cells. In addition, unlike the result with normal BALB/c mice, DBA/2-stimulated CTL do not react with B10.D2 targets (Figs 3, 4) or DBA/1 targets (Fig. 4) but react with (DBA/1 $\times$ BALB/c)_{F1} targets.

We interpret these results as indicating that in the thymus-grafted mice, H-2^d-restricted CTL-P specific for B6 minor histocompatibility antigens, including those specific for minor histocompatibility antigens shared by DBA/2 and B6, are tolerized, while those specific for DBA/2 minor histocompatibility antigens not shared with B10.D2 are not tolerized. An analogous result was obtained with cells from a BALB/c *nu/nu* mouse grafted with B10.BR 2-deoxyguanosine-treated thymuses; DBA/2 induced H-2^d-restricted DBA minor histocompatibility antigen-specific CTL which did not cross-react on B10.D2 targets. B10.D2 cells induced no CTL. Thus, in all *nu/nu* mice grafted with either B6 or B10.BR thymus, T cells were tolerant to donor minor histocompatibility antigens in the context of recipient-type MHC antigens but were reactive to donor-type major histocompatibility antigens.

It is unlikely that the minor histocompatibility antigens presented directly by the epithelial cells tolerize the T cell, as class I MHC antigens expressed by the epithelial cells do not induce tolerance. The experimental situation reported here shows several differences from a situation in which thymic minor

Fig. 4 Six-week-old BALB/c *nu/nu* mice were transplanted with either B6 or B10.BR 2-deoxyguanosine-treated thymuses as described in Fig. 1 legend. Eight weeks later the mice were primed by injecting i.p. 10^7 X-irradiated (2,200R) spleen cells each of DBA/2 and B10.D2 mice. Two weeks later spleen cells were stimulated with various stimulators as indicated. All cultures except those containing B6 and B10.BR stimulators contained IL-2⁹. Cytolytic activity was tested after 5 days of culture⁹.

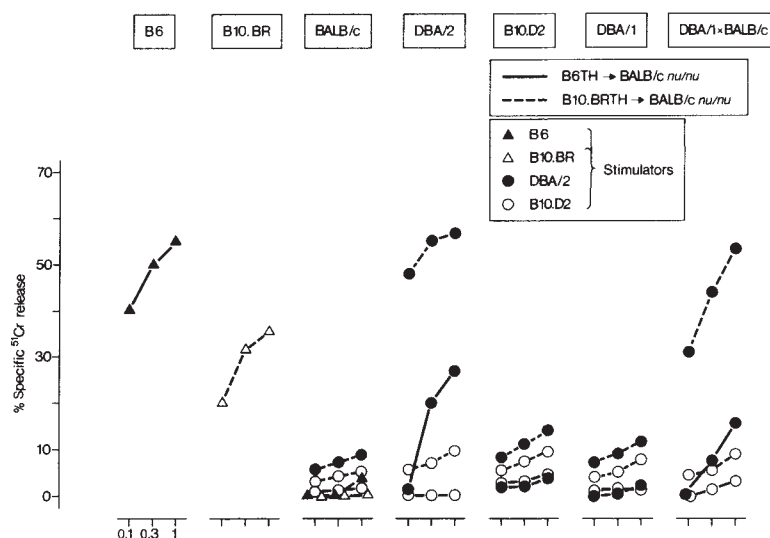


Table 1 Response in mixed lymphocyte culture of thymocytes from *nu/nu* mice transplanted with 2-deoxyguanosine-treated thymuses

Responders (5×10^5 thymocytes from)	Stimulator (10^6) spleen cells			
	BALB/c	B6	CBA/J	B10.BR
BALB/c	0.350	10.500	15.500	—
B6	6.000	1.100	6.300	—
BALB/c $\leftarrow$ B6TH	0.600	4.000	—	8.600
BALB/c $\leftarrow$ B6TH	0.500	3.800	8.000	—
BALB/c $\leftarrow$ B6TH	1.400	6.900	7.400	—
BALB/c $\leftarrow$ B10.BR TH	1.100	14.000	—	3.400
BALB/c $\leftarrow$ B10.BR TH	1.100	9.700	—	4.700

Different BALB/c *nu/nu* mice were grafted with three lobes per kidney of 14-day-old fetal B6 thymuses treated *in vitro* with 2-deoxyguanosine. Stimulator cells were treated with anti-Thy-1 serum plus complement and X-irradiated (2,000R).

antigens did not tolerize CTL-P in the context of recipient MHC antigens¹¹. In these experiments animals were transplanted with X-irradiated thymuses. One might argue that in this situation donor-type thymic antigen-presenting cells were not replaced by host cells and therefore no tolerance to donor minor histocompatibility antigens in the context of host-type major histocompatibility antigens was induced. Since these animals were

tolerant to host-type antigens, we must conclude that tolerance can also be induced by cells unable to present antigens derived from other cells in tolerogenic form.

Our results indicate that haematopoietic cells play an important part in the tolerization of MHC antigen-specific T lymphocytes² and indicate that haematopoietic cells are able to present minor but not major epithelial cell-derived histocompatibility antigens in tolerogenic form. The existence of this type of antigen-presenting cell provides a satisfactory explanation for the finding of apparently unrestricted tolerance to male antigen¹².

We thank Drs W. Haas and P. Matzinger for discussion and Drs W. Haas and R. Scollay for reading the manuscript.

Received 22 January; accepted 28 February 1986.

- Jenkinson, E. J., Franchi, L., Kingston, R. & Owen, J. J. T. *Eur. J. Immun.* **12**, 583–587 (1982).
- von Boehmer, H. & Schubiger, K. *Eur. J. Immun.* **14**, 1048–1052 (1984).
- Jordan, R. K., Robinson, J. H., Hopkinson, N. A., House, K. C. & Bentley, A. C. *Nature* **314**, 454–456 (1985).
- Jenkinson, E. J., Jhittay, P., Kingston, R. & Owen, J. J. T. *Transplantation* **30**, 331–333 (1985).
- Summerlin, W. T. *Clin. Immun. Immunopath.* **1**, 372–374 (1973).
- Lafferty, K. J., Books, A., Dart, G. & Talmage, D. N. *Transplantation* **225**, 138–141 (1976).
- Bevan, M. J. *J. exp. Med.* **143**, 1283–1288 (1976).
- Gordon, R. D., Nathieson, B. J., Samelson, L. E., Boyse, E. A. & Simpson, E. J. *exp. Med.* **144**, 810–820 (1976).
- Teh, H. S., Bennick, J. & von Boehmer, H. *Eur. J. Immun.* **12**, 887–892 (1982).
- Fischer Lindahl, K. & Langhorne, J. *Scand. J. Immun.* **14**, 643–654 (1981).
- Matzinger, P., Zamyoska, R. & Waldmann, H. *Nature* **308**, 738–741 (1984).
- von Boehmer, H., Haas, W. & Pohlitz, H. J. *exp. Med.* **147**, 1291–1295 (1978).

Concerted activation of two potential proto-oncogenes in carcinomas induced by mouse mammary tumour virus

Gordon Peters*, Audrey E. Lee† & Clive Dickson†

* Imperial Cancer Research Fund Laboratories, St Bartholomew's Hospital, London EC1A 7BE, UK

† Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX, UK

The induction of tumours by retroviruses lacking transduced oncogenes can involve the transcriptional or functional activation of cellular proto-oncogenes by an integrated provirus^{1–8}. Thus, the two cellular genes *int-1* and *int-2*, identified as common targets for activation by mouse mammary tumour virus (MMTV), may constitute previously unrecognized oncogenes^{9,10}. In tumours, proviral insertion at these loci leads to expression of messenger RNAs which are undetectable in normal mammary glands^{11,12}. Here we report that in a survey of the two transcriptional activity and structural integrity of the two *int* loci in 30 BR6 mouse mammary tumours, around 50% of the tumours expressed both of these genes, in ostensibly monoclonal cell populations. Our data suggest that *int-1* and *int-2* may act cooperatively in the genesis of mammary carcinomas. However, because three tumours (10%) involved neither gene, and because in five cases activation occurred in the apparent absence of an adjacent provirus, it is clear that other loci and mechanisms contribute to tumorigenesis.

DNA of high relative molecular mass (M_r) and total cellular RNA were prepared from a randomly selected series of mammary tumours which arose spontaneously in BR6 mice¹³. Expression of the *int-1* gene was assessed using a 2.5-kilobase (kb) *EcoRI/BamHI* fragment (probe C in Fig. 2a, provided by R. Nusse)^{11,14} to probe blots of polyadenylated RNA. The major transcript detected was the expected 2.6-kb *int-1* RNA¹¹ (Fig. 1a), although in some instances (such as tumour F5) abnormally sized transcripts were apparent, resulting from provirus insertion within known exons of the *int-1* gene^{11,14}. Similar analyses were performed for *int-2* RNA expression. Using a 532-base-pair

EcoRI/SacI fragment (designated *int-2-f* in Fig. 2a) as a hybridization probe¹², two major *int-2* RNAs of approximately 3.2 and 3.0 kb were detected (Fig. 1b). We are now investigating the precise structure of these two transcripts, as we previously detected only the 3.2-kb RNA using a different probe, *int-2-c* (ref. 12).

Of the 30 tumours examined, 27 expressed *int-1* or *int-2* RNA (Table 1). However, the most striking result was that almost 50% of the tumours expressed both RNAs at comparable levels. Only three tumours failed to express either *int-1* or *int-2* (see Table 1), or did so at levels which precluded detection by our methods. As their existence implied that other genes or mechanisms must be involved in tumorigenesis in such cases, it was important to confirm that the *int-1* and *int-2* genes were unaltered in these three tumours. Moreover, because of the frequency of expression of the two genes, it was essential to establish whether they were activated within the same cell. MMTV-induced tumours are generally clonal or quasi-clonal, as judged by the relative intensities of acquired (haploid) versus endogenous (diploid) viral sequences or rearranged versus normal alleles in DNA blotting analyses^{9,10,15}. Southern blots were therefore performed on all 30 tumour DNAs using a panel of restriction enzymes which afforded good resolution and range with the probes available. The use of five enzymes with either the A or C probes from *int-1* permitted visualization of DNA fragments spanning over 30 kb of chromosomal DNA (Fig. 2a and refs 11, 14). Similarly, blots hybridized with the *int-2* probes a', c, f and j covered over 40 kb of DNA (Fig. 2a and ref. 12). With most of the tumours (see, for example, E155A in Fig. 2b), DNA rearrangements were detected with multiple enzyme probe combinations, and as we had chosen enzymes that cleaved MMTV proviral DNA at known sites, it was possible to deduce both the position and transcriptional orientation of each provirus (Fig. 2a).

As predicted by previous reports on both *int-1* and *int-2*^{11,12}, proviruses located 5' of each gene were generally in the opposite transcriptional orientation while those on the 3' side were in the same orientation as the gene. However, tumour Z79 showed a provirus in *int-1* in a promoter insertion mode, within the first exon, and a similar interpretation has been placed on the *int-2* insertions in tumours D157B and W26. Curiously, tumour U153 showed a provirus 3' of *int-1* but in an opposite orientation,

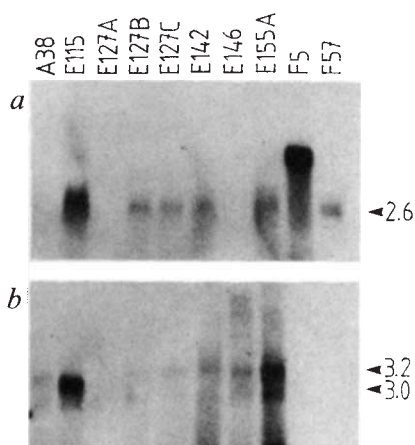


Fig. 1 Analysis of *int-1* and *int-2* RNAs in mouse mammary tumours. RNA from each of the 10 mammary tumours indicated was analysed by Northern blotting and hybridization with the *int-1*-C probe (a) and the *int-2*-f probe (b) (see Fig. 2a for probe designation). Numbers on the right indicate the sizes (kb) and positions of the major RNA transcripts.

Methods. Polyadenylated RNA was prepared from 30 BR6 mouse mammary tumours by extraction with guanidine thiocyanate and chromatography on poly(rU)-Sephadex¹². Aliquots (5 µg) were fractionated by electrophoresis in 1.2% agarose gels containing 2.2 M formaldehyde and transferred to nitrocellulose by blotting in 20×SSC¹². Hybridizations were performed at 50 °C in 50% formamide as described elsewhere and filters were washed in 0.1×SSC, 0.1% SDS at 65 °C before autoradiography¹⁰. The probes used were a 2.5-kb *EcoRI*/*Bam*HI fragment from *int-1*, designated probe C by Nusse *et al.*^{11,14} (a), and a 0.5-kb *EcoRI*/*Sac*I fragment from *int-2*, probe f¹² (b), nick-translated with [α -³²P]dCTP¹⁰ to a specific activity of 10⁸ c.p.m. µg⁻¹. The sizes of the indicated RNAs were assessed relative to 28S and 18S ribosomal RNAs, and to the MMTV 35S and 24S mRNA species detected by hybridizing the same blots with an MMTV long terminal repeat (LTR) probe (not shown). This also confirmed the integrity of RNA samples which scored negative with *int-1* or *int-2* probes.

contradicting the accepted dogma of enhancer-mediated activation of these genes^{11,12,15}

Table 1 summarizes the conclusions drawn from these Southern blotting data. First, detection of MMTV proviral sequences within the *int-1* or *int-2* domains invariably correlated with detection of the respective RNA. Surprisingly, the converse was not always true, in that five tumours expressed *int-1* or *int-2* RNA in the apparent absence of an activating provirus or discontinuity in the DNA. However, the three tumours scoring negative for both RNAs were negative for proviral insertions in both *int* loci, although they contained multiple MMTV proviruses in other chromosomal sites and expressed viral RNA (not shown). Thus, although viral aetiology and apparent clonality could be established in these and subsequently all the tumours analysed, mechanisms other than proviral activation of *int-1* and *int-2* may contribute to tumorigenesis in such cases.

The example shown for tumour E155A in Fig. 2b was typical of the data obtained in these analyses in that the intensities of the rearranged DNA sequences were approximately equivalent to those from the normal alleles. As this was true for both the *int-1* and *int-2* probes, our data imply that E155A is an example of a monoclonal tumour in which both genes are activated. Analogous results were obtained for the other tumours which showed evidence for insertion within both *int* loci, although there were rare cases of oligo-clonality.

In a further attempt to establish the clonal nature of the tumours, a number of them were dissociated and transplanted into multiple BR6 mice. RNA and DNA were then prepared from independent transplants and analysed by Northern and Southern blot hybridization. All the transplanted tumours retained the capacity to express both *int-1* and *int-2* RNAs at

Table 1 Involvement of *int-1* and *int-2* in BR6 mouse mammary tumours

Tumour	Expression of RNA		Presence of MMTV provirus	
	<i>int-1</i>	<i>int-2</i>	<i>int-1</i>	<i>int-2</i>
D124	+	+	+	+
E115	+	+	+	+
E142A	+	+	+	+
E155A	+	+	+	+
T108	+	+	+	+
U5F	+	+	+	+
U153	+	+	+	+
W16B	+	+	+	+
W106	+	+	+	+
Z79	+	+	+	+
A38	+	+	+	—
V97B	+	+	+	—
W46A	+	+	+	—
E127C	+	+	—	+
D157A	+	—	+	—
F5	+	—	+	—
F57	+	—	+	—
U79	+	—	+	—
W8	+	—	+	—
X98	+	—	+	—
X123	+	—	+	—
C46A	—	+	—	+
D157B	—	+	—	+
E146	—	+	—	+
V97A	—	+	—	+
W26	—	+	—	+
E127B	—	+	—	—
E127A	—	—	—	—
S40	—	—	—	—
T103	—	—	—	—

The data shown are exemplified by Figs 1 and 2 in which the structural integrity and transcriptional activity of the *int-1* and *int-2* loci were assessed by blot hybridization of DNA and RNA from 30 spontaneous tumours. Individual mice are identified by a letter and number code and where this is followed by an additional letter, several independent tumours were recovered from the same animal. The tumours are divided into four groups based on detection of *int-1* or *int-2* RNAs, as depicted by + or —. The presence or absence of MMTV proviral sequences within the 30 kb surrounding each locus is also indicated (+/—).

levels comparable to those found in the original tumour (data not shown). More significantly, the various transplants did not segregate cell clones that differed in their pattern of *int-1*, *int-2* or MMTV-specific DNA fragments. Figure 3 shows an example for the primary tumour E115, which contained four acquired MMTV proviruses. These were distinguishable as 5'- or 3'-specific *Eco*RI fragments at about half the intensity of equivalent junction fragments derived from the endogenous MMTV sequences of BR6 mice (Fig. 3a,b). As confirmed using *Bgl*II-digested DNA, two of these proviruses had disrupted the *int-1* (Fig. 3c) and *int-2* (Fig. 3d) loci (see Fig. 2a). Equivalent analyses on four of the transplanted tumours derived from E115 indicated that they all retained the same patterns and intensities of restriction fragments and had therefore not segregated independent clones. Moreover, one of the transplanted tumours from E115 (tumour 2) had acquired an additional MMTV provirus, presumably as a result of superinfection (Fig. 3a,b). As the intensities of the novel fragments generated from this provirus suggested that it was present in all cells of the tumour, these data imply that transplant 2 must have become established from a single or, at most, very few E115 tumour cells, and confirms that activation of both the *int-1* and *int-2* genes was occurring within the same cell.

The results summarized in Table 1 suggest that concerted activation of the two *int* genes within the same cell is not a

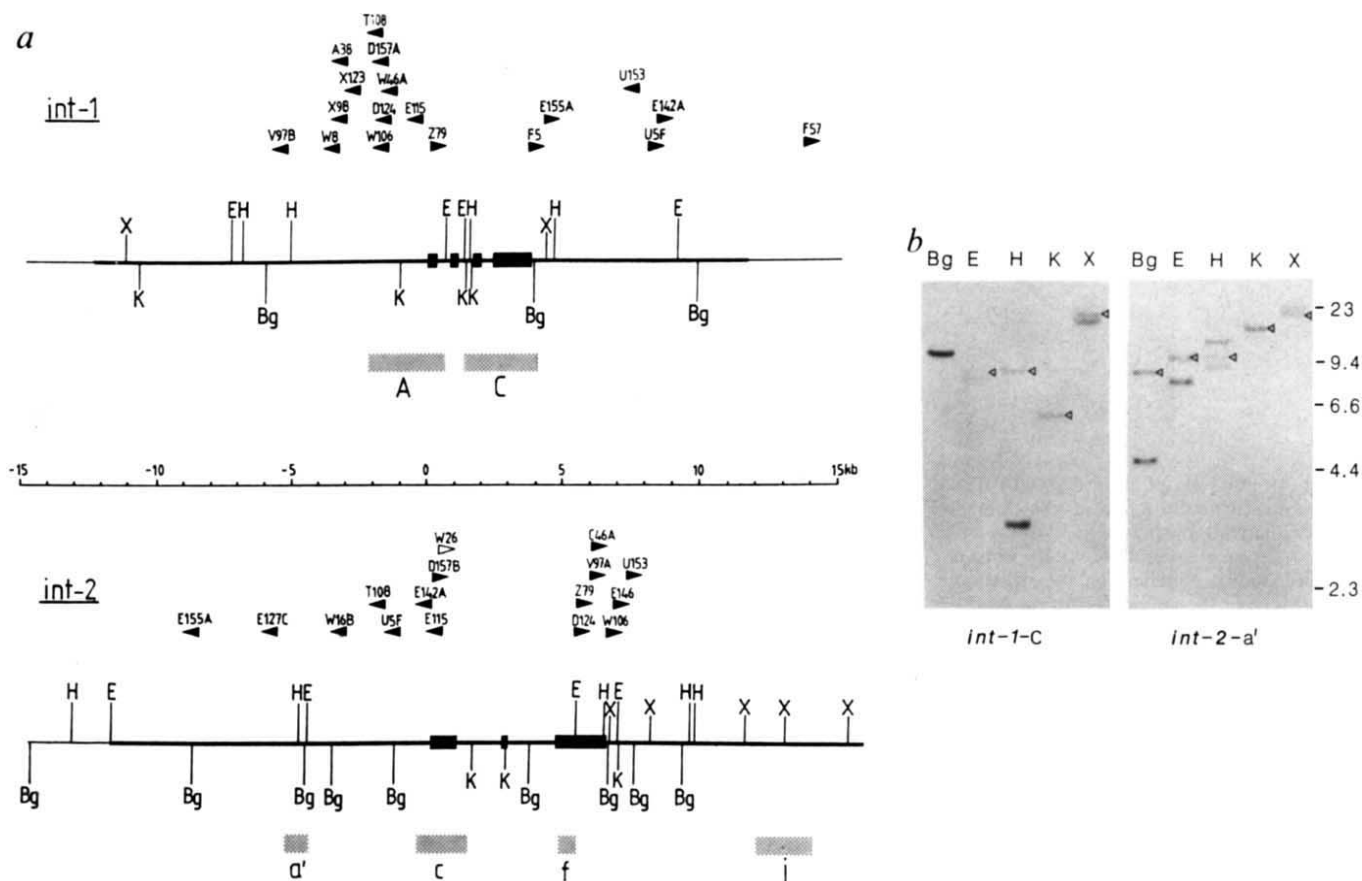


Fig. 2 Topography of *int-1* and *int-2* and sites of proviral insertion. *a*, The sites of cleavage for the enzymes *Bgl*II (Bg), *Eco*RI (E), *Hind*III (H), *Kpn*I (K) and *Xba*I (X) in DNA spanning over 30 kb of the *int-1* and *int-2* loci^{11,12}. The maps are aligned so that the presumed 5' end of each gene is at map position 0. The approximate extents of exons, taken from the published data of van Ooyen and Nusse for *int-1*¹⁴, and from observations on *int-2* cDNAs (R. Moore, personal communication, and ref. 32), are depicted as solid boxes. Arrowheads above each map signify the integration sites and transcriptional orientation of MMTV proviruses detected within these regions in the numbered tumours. The open arrowhead identifies a solo MMTV LTR in tumour W26 (ref. 32). The extents and designation of restriction fragments selected as specific probes are shown as stippled bars beneath each map. Note that the *int-2* map has been extended to include an additional 7.8-kb *Eco*RI fragment at the 5' end (M. Placzek, personal communication) and is shown in a different orientation from that reported previously to take account of our revised notion of transcriptional direction³². *b*, Southern blots of DNA from a single mammary tumour (E155A) digested with the five restriction enzymes whose cleavage sites are indicated in *a*, and hybridized to either the *int-1-C* or *int-2-a'* probe. $\triangleleft$, Novel fragments generated by insertion of an MMTV provirus.

Methods. High-M, DNA was prepared from BR6 mammary tumours following standard protocols¹⁰. Aliquots (10 μ g) were digested with the indicated enzymes and fractionated by electrophoresis through 0.8% agarose gels¹⁰. Following denaturation and transfer to nitrocellulose, the DNAs were hybridized sequentially with a series of cloned DNA probes specific for the *int-1* and *int-2* loci as depicted in *a*. The cloned fragments, A and C from *int-1*^{11,14}, and a', c, f and j from *int-2*¹², were excised from their respective plasmid vectors and labelled by nick-translation with [α -³²P]dCTP¹⁰. Hybridizations were performed at 42 °C in 50% formamide as described previously¹⁰, and filters were washed at 65 °C in 0.1 $\times$ SSC, 0.1% SDS. To perform sequential hybridization on the same blots, the nitrocellulose filters were rinsed in 30 mM NaOH for 5 min at room temperature to remove the ³²P-labelled probes. Numbers on the right identify the size (kb) and position of fragments of *Hind*III-digested λ DNA used as markers in these analyses.

particularly rare event, for we observed 10 examples (33%) in which RNA expression could be attributed directly to integration of a MMTV provirus in both domains. Since each represents about 10⁻⁵ of the total mouse genome and as integration is considered to be random¹⁵, the chances of coincidentally finding a cell clone in which both regions are perturbed by a provirus are small, unless their dual activation confers additional growth advantages on that cell.

Cooperativity between potential oncogenes has various precedents, as exemplified by transfection experiments *in vitro*¹⁶⁻¹⁸ and by the transduction of segments from two discrete genes by several acutely oncogenic retroviruses¹⁹⁻²⁷. Since transduction of a cellular gene is thought to be preceded by provirus integration in adjacent DNA, it is conceivable that the generation of such dual oncogene retroviruses occurred in cells in which both genes had been targets for provirus insertion. The examples described here, the concerted activation of the *pim-1* and *myc* genes in murine thymic lymphomas²⁸ and the double insertions within *Mlvi-1* and *Mlvi-2* in rat thymomas^{29,30} may

therefore reflect analogous situations which do not progress to transduction.

Since proviral activation of two discrete genes occurs so frequently, it is possible that expression of either one alone is insufficient for overt malignancy. In the MMTV system, this notion is supported by our previous studies on pregnancy-dependent mammary tumours, which implicate the *int-2* gene³¹. Because these tumours regress between pregnancies, we argued that some additional event or stimulus was required for their subsequent transition to hormone-independence and true malignancy. Proviral insertion within *int-1* may represent one such event and it will be of interest to determine when and in what order *int-1* and *int-2* activation occurs during the progression of these recurring tumours. However, there are clearly many other genes which may be involved in these processes, possibly acting in concert with either *int-1* or *int-2*. It is therefore tempting to invoke such genes in tumours in which we detected only one of the *int* RNAs. Whether they are also activated by proviruses or by some completely independent mechanism is unknown,

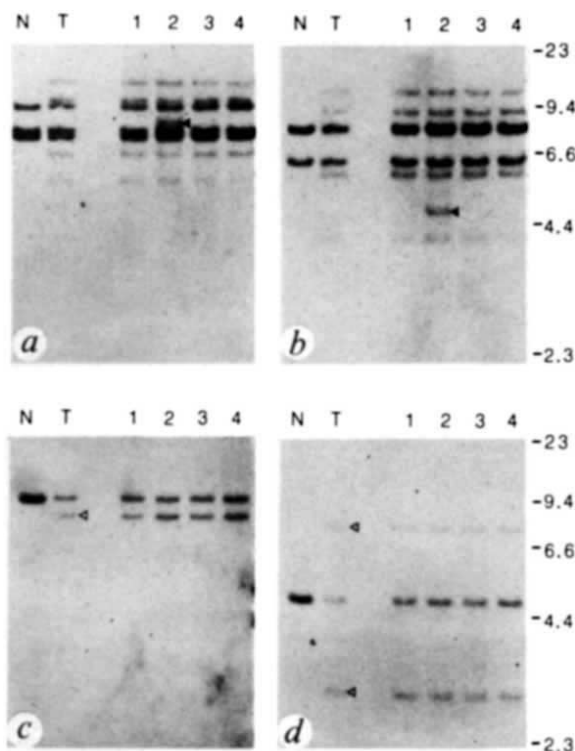


Fig. 3 Clonality of primary and transplanted mouse mammary tumours. DNAs from normal tissue (N), from the primary BR6 mammary tumour E115 (T) and from a series of four transplanted tumours derived from E115 (1–4) were analysed by Southern blotting. *a*, *EcoRI*-digested DNA probed with MMTV 5'-specific sequences¹⁰; *b*, *EcoRI*-digested DNA probed with MMTV 3'-specific sequences; *c*, *BglII*-digested DNA probed with *int-1*-C; *d*, *BglII*-digested DNA probed with *int-2*-c. Since *EcoRI* cleaves the MMTV provirus at a single site, each band in the autoradiographs in *a* and *b* represents a unique junction between viral and cellular DNA sequences^{9,10}. The positions of novel restriction fragments resulting from proviral insertion within *int-1* and *int-2* are indicated (<). ◀ Denotes a new pair of *EcoRI* fragments presumably resulting from superinfection of tumour 2 by MMTV. **Methods.** Segments of primary mammary tumours which had been stored frozen at -70°C were disrupted into a rough cell suspension and injected subcutaneously into a series of female BR6 mice. Each animal received $\sim 10^4$ cells, but since no particular measures were taken to ensure the viability of the frozen tissues, we estimate that only a small percentage of these cells were viable. On average, tumours arose from these transplanted cells within 6 weeks and were generally excised when they had grown to $\sim 1\text{ cm}^3$. High- M_r DNA and total cellular RNAs were prepared from the original and transplanted tumours and analysed by gel electrophoresis and blot hybridization as described for Figs 1 and 2.

but the 10% of MMTV-induced tumours in which *int-1* and *int-2* are not expressed strongly suggest that a further search for 'int-like' oncogenes may prove rewarding.

We thank J. Witkowski, J. Wyke, R. Moore and M. Placzek for useful discussions and comments on the manuscript, and A. Sterlini for help in its preparation. The *int-1* probes were kindly provided by R. Nusse of the Netherlands Cancer Institute. Much of this work was carried out in the Vera Simpson Memorial Laboratory at St Bartholomew's Hospital.

Received 16 December 1985; accepted 6 March 1986.

- Hayward, W. S., Neel, B. G. & Astrin, S. M. *Nature* **290**, 475–480 (1981).
- Noori-Daloii, M. R. *et al. Nature* **294**, 574–576 (1981).
- Fung, Y. K. T., Lewis, W. G., Crittenden, L. B. & Kung, H.-J. *Cell* **33**, 357–368 (1983).
- Neil, J. C. *et al. Nature* **308**, 814–820 (1983).
- Corcoran, L. M., Adams, J. M., Dunn, A. R. & Cory, S. *Cell* **37**, 113–122 (1984).
- Li, Y., Holland, C. A., Hartley, J. W. & Hopkins, N. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6808–6811 (1984).
- Steffen, D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2097–2101 (1984).
- Shen-Ong, G. L., Potter, M., Mushinski, J. F., Lavu, S. & Reddy, E. P. *Science* **226**, 1077–1080 (1984).

- Nusse, R. & Varmus, H. E. *Cell* **31**, 99–109 (1982).
- Peters, G., Brookes, S., Smith, R. & Dickson, C. *Cell* **33**, 364–377 (1983).
- Nusse, R., Van Ooyen, A., Cox, D., Fung, Y. K. T. & Varmus, H. *Nature* **307**, 131–136 (1984).
- Dickson, C., Smith, R., Brookes, S. & Peters, G. *Cell* **37**, 529–536 (1984).
- Foulds, L. *Br. J. Cancer* **3**, 230–239 (1949).
- Van Ooyen, A. & Nusse, R. *Cell* **39**, 233–240 (1984).
- Varmus, H. E. *Cancer Surv.* **1**, 309–319 (1982).
- Cooper, G. M. & Neiman, P. E. *Nature* **292**, 857–858 (1981).
- Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596–602 (1983).
- Ruley, H. E. *Nature* **304**, 602–606 (1983).
- Vennstrom, B. & Bishop, J. M. *Cell* **23**, 135–143 (1982).
- Frykberg, L. *et al. Cell* **32**, 227–238 (1983).
- Coll, J. *et al. EMBO J.* **2**, 2189–2194 (1983).
- Jansen, H. W., Ruckert, B., Lurz, R. & Bister, K. *EMBO J.* **2**, 1969–1975 (1983).
- Kan, N. C., Flordellis, C. S., Garon, C. F., Duesberg, P. H. & Papas, T. S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6566–6570 (1983).
- Van Beveren, C., van Straaten, F., Curran, T., Muller, R. & Verma, I. M. *Cell* **32**, 1241–1255 (1983).
- Nunn, M. F., Seeburg, P. H., Moscovici, C. & Duesberg, P. H. *Nature* **306**, 391–395 (1983).
- Leprince, D. *et al. Nature* **306**, 395–397 (1983).
- Naharro, G., Robbins, K. C. & Reddy, E. P. *Science* **223**, 63–66 (1984).
- Selten, G., Cuyper, H. T., Zijlstra, M., Melief, C. & Berns, A. *EMBO J.* **3**, 3215–3222 (1984).
- Tsichlis, P. N., Strauss, P. G. & Hu, L.-F. *Nature* **302**, 445–449 (1983).
- Tsichlis, P. N., Strauss, P. G. & Lohse, M. A. *J. Virol.* **56**, 258–267 (1985).
- Peters, G., Lee, A. E. & Dickson, C. *Nature* **309**, 273–275 (1984).
- Moore, R. *et al. EMBO J.* (in the press).

The inositol tris/tetrakisphosphate pathway—demonstration of $\text{Ins}(1,4,5)\text{P}_3$ 3-kinase activity in animal tissues

R. F. Irvine*, A. J. Letcher*, J. P. Heslop† & M. J. Berridge†

* Department of Biochemistry, AFRC Institute of Animal Physiology, Babraham, Cambridge CB2 4AT, UK

† AFRC Unit of Insect Neurophysiology and Pharmacology, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Recent advances in our understanding of the role of inositides in cell signalling have led to the central hypothesis that a receptor-stimulated phosphodiesteratic hydrolysis of phosphatidylinositol 4,5-bisphosphate ($\text{PtdIns}(4,5)\text{P}_2$)^{1,2} results in the formation of two second messengers, diacylglycerol³ and inositol 1,4,5-trisphosphate ($\text{Ins}(1,4,5)\text{P}_3$)⁴. The existence of another pathway of inositide metabolism was first suggested by the discovery that a novel inositol trisphosphate, $\text{Ins}(1,3,4)\text{P}_3$, is formed in stimulated tissues⁵; the metabolic kinetics of $\text{Ins}(1,3,4)\text{P}_3$ are entirely different from those of $\text{Ins}(1,4,5)\text{P}_3$ (refs 6, 7). The probable route of formation of $\text{Ins}(1,3,4)\text{P}_3$ was recently shown to be via a 5-dephosphorylation of inositol 1,3,4,5-tetrakisphosphate ($\text{Ins}(1,3,4,5)\text{P}_4$), a compound which is rapidly formed on muscarinic stimulation of brain slices, and which can be readily converted to $\text{Ins}(1,3,4)\text{P}_3$ by a 5-phosphatase in red blood cell membranes⁸. However, the source of $\text{Ins}(1,3,4,5)\text{P}_4$ is unclear, and an attempt to detect a possible parent lipid, phosphatidylinositol 3,4,5-trisphosphate ($\text{PtdIns}(3,4,5)\text{P}_3$), was unsuccessful⁹. The recent discovery that the higher phosphorylated forms of inositol (InsP_5 and InsP_6) also exist in animal cells⁹ suggested that inositol phosphate kinases might not be confined to plant and avian tissues, and here we show that a variety of animal tissues contain an active and specific $\text{Ins}(1,4,5)\text{P}_3$ 3-kinase. We therefore suggest that an inositol tris/tetrakisphosphate pathway exists as an alternative route to the dephosphorylation^{10,11} of $\text{Ins}(1,4,5)\text{P}_3$. The function of this novel pathway is unknown.

The kinase was studied in both intact and broken cell preparations. When ^3H -labelled *myo*- $\text{Ins}(1,4,5)\text{P}_3$ was microinjected into *Xenopus* oocytes (Fig. 1), most was dephosphorylated into inositol and the lower inositol phosphates, but a considerable proportion (10–20%) was rapidly converted into a compound which, on HPLC ion-exchange chromatography⁹, ran as an inositol tetrakisphosphate. Injection of ^3H -*myo*-inositol far in excess of that derived from dephosphorylation of ^3H - $\text{Ins}(1,4,5)\text{P}_3$ resulted in no such InsP_4 production, indicating that the incorporation of radiolabel into InsP_4 was not caused

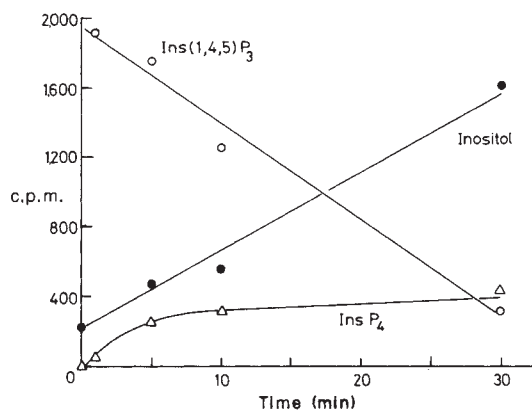


Fig. 1 Metabolism of $\text{Ins}(1,4,5)\text{P}_3$ by intact *Xenopus* oocytes. Individual oocytes were injected with $^3\text{H}\text{Ins}(1,4,5)\text{P}_3$ and after various incubation periods labelled compounds were extracted and separated by HPLC. Apart from the original $^3\text{H}\text{-myo-Ins}(1,4,5)\text{P}_3$ ($\circ$), the major metabolic products were ^3H -inositol ($\bullet$) and $^3\text{H}\text{-Ins}(1,3,4,5)\text{P}_4$ (Δ). Trace amounts of ^3H -inositol monophosphate ($^3\text{H}\text{-Ins}(1)\text{P}$) and $^3\text{H}\text{-Ins}(1,4)\text{P}_2$ were detected in some of the extracts, but the counts were too low for precise quantification. The data points represent single oocytes; similar results were obtained in an identical experiment.

Methods. A 10- μl aliquot of $^3\text{H}\text{-myo-Ins}(1,4,5)\text{P}_3$ (specific activity 1 Ci mmol^{-1} ; Amersham) was dried under nitrogen in the bottom of a Perspex tube. The dried residue was redissolved in 200 μl of an injection solution containing NaCl (70 mM) and Tris-HCl (10 mM, pH 7.2). Immature oocytes were injected with 7 μl of solution containing $\sim 2,000$ c.p.m. of $^3\text{H}\text{-Ins}(1,4,5)\text{P}_3$. After injection each oocyte was incubated for various times in 400 μl of modified Barth's saline (88 mM NaCl, 2 mM KCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 0.5 mM Na_2HPO_4 , 2 mM NaHCO_3 , 15 mM HEPES (pH 7.6). This saline was removed before the addition of 200 μl of 5% trichloroacetic acid (TCA). After addition of 500 μl of distilled water, each extract was applied without filtration to a Partisil 10 SAX anion-exchange HPLC column and the individual inositol phosphates were separated as described in ref. 9. In a separate series of experiments oocytes were injected with 7 μl of a solution containing 52,000 c.p.m. of $\text{myo-[2-}^3\text{H]inositol}$ (Amersham). After a 30-min incubation, label was extracted and run on HPLC as described above. There was a small amount of $^3\text{H}\text{-Ins}(1)\text{P}$ but no accumulation of label corresponding to $\text{Ins}(1,4)\text{P}_2$, $\text{Ins}(1,4,5)\text{P}_3$ or $\text{Ins}(1,3,4,5)\text{P}_4$.

by recycling of liberated $^3\text{H}\text{-myo-inositol}$ through the inositol lipids. No accumulation of $\text{Ins}(1,3,4)\text{P}_3$ was detected in these injection experiments, which is probably a result of the relative rates of dephosphorylation of InsP_4 and $\text{Ins}(1,3,4)\text{P}_3$ in these cells, and also of the low number of c.p.m. injected (that is, if $\text{Ins}(1,3,4)\text{P}_3$ had constituted $<10\%$ of the InsP_4 radioactivity, we would not have detected it).

Having established the ability of an intact cell to phosphorylate $\text{Ins}(1,4,5)\text{P}_3$, kinase activity was studied further in broken cell preparations. In the three rat tissues examined (brain, liver and pancreas), $\text{Ins}(1,4,5)\text{P}_3$ was rapidly phosphorylated to InsP_4 . Under appropriate conditions (see Fig. 2 legend), 60–70% of $\text{Ins}(1,4,5)\text{P}_3$ was converted by a 2% brain homogenate into InsP_4 in 5 min, with the consequent formation of some $\text{Ins}(1,3,4)\text{P}_3$ (Fig. 2). Using these conditions, a bulk preparation of pure, de-salted InsP_4 was made⁸, and the experiments used previously to establish its structure in carbachol-stimulated brain slices⁸ were repeated. The compound was identified as an InsP_4 by iontophoresis and paper chromatography with internal standards. The distribution of the phosphates in the inositol ring was determined by dephosphorylation using calf intestine alkaline phosphatase and human red cell ghosts, yielding two forms of InsP_3 which, after periodate oxidation, borohydride reduction and dephosphorylation, yielded only xylitol and altritol respectively as products. (More detailed descriptions of the techniques and the logic of these experiments are given in refs 5 and 8.)

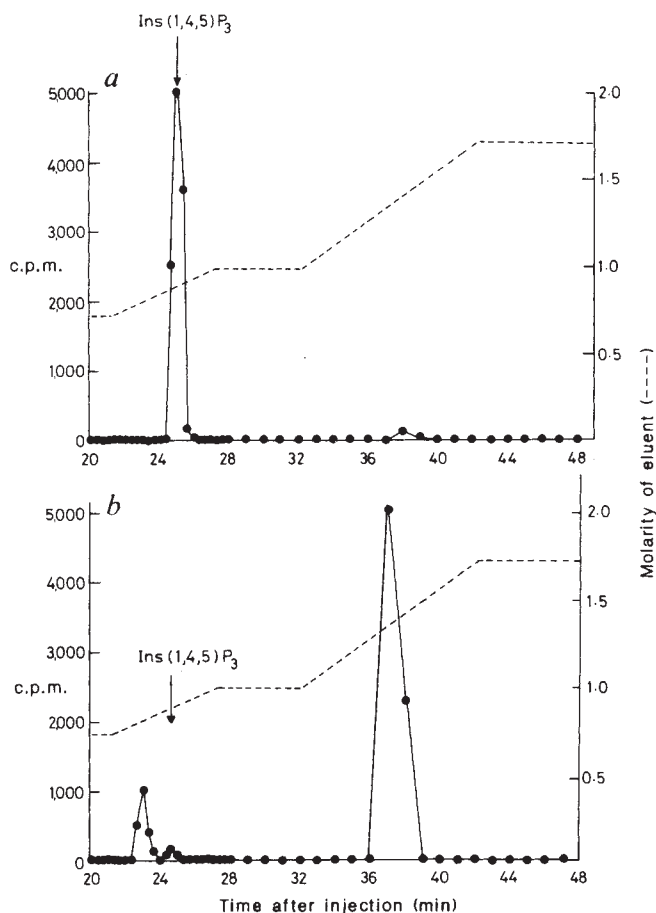


Fig. 2 Formation of $\text{Ins}(1,3,4,5)\text{P}_4$ and $\text{Ins}(1,3,4)\text{P}_3$ in a broken cell preparation. 10 pmol of $^3\text{H}\text{-myo-Ins}(1,4,5)\text{P}_3$ (specific activity 1 Ci mmol^{-1} ; Amersham) were incubated for 5 min at pH 7.5 (0.05 M Tris-maleate) with 0.05 ml of a 2% homogenate (in 0.15 M sucrose) of rat brain, in a final volume of 0.2 ml containing 10 mM ATP 20 mM MgCl_2 . The reaction was quenched with 0.5 ml of 20% ice-cold TCA and 0.05 ml of 5% bovine serum albumin (BSA), then the TCA was removed by four diethylether washes and the mixture neutralized with NH_3 (ref. 22). The resulting solution (2 ml) was loaded onto a Partisil SAX anion-exchange HPLC column and fractionated as described in ref. 8 (b). The original $^3\text{H}\text{-Ins}(1,4,5)\text{P}_3$ was analysed using the same technique (a). Note the appearance of InsP_4 (38 mins) and of another InsP_3 (23 mins) which is probably $\text{Ins}(1,3,4)\text{P}_3$ (refs 6–8). The gradient of ammonium formate, buffered to pH 3.7 with phosphoric acid, is superimposed on the elution profiles. The position of an internal $^{32}\text{P}\text{-Ins}(1,4,5)\text{P}_3$ marker¹⁰ is indicated by the arrows. Analysis of part of the sample shown in b by Dowex anion-exchange chromatography showed that $\sim 15\%$ of the radioactivity had been converted to inositol, InsP or InsP_2 ; only the InsP_3 and InsP_4 regions of the HPLC separation were collected. For routine assays to characterize the $\text{Ins}(1,4,5)\text{P}_3$ kinase, the same techniques were used, but analysis of InsP_4 radioactivity was by means of standard Dowex anion-exchange chromatography⁸. These experiments were repeated using $^{32}\text{P}\text{-Ins}(1,2\text{-cyclic-}4,5)\text{P}_3$; we identified this on our HPLC separations as the earlier-eluting radioactive compound released from red blood cell ghosts on incubation with Ca^{2+} (ref. 6). On HPLC⁶, iontophoresis at pH 3.6 (ref. 6) or paper chromatography⁵, migration of $^{32}\text{P}\text{-Ins}(1,2\text{-cyclic-}4,5)\text{P}_3$ is distinct from that of $\text{Ins}(1,4,5)\text{P}_3$, but after mild treatment with acid (0.1 M HCl, 50 °C for 10 min or 20% TCA, 30 min at 0 °C) the former migrates in all these separation procedures with internal $\text{Ins}(1,4,5)\text{P}_3$ markers. On the basis of this result and data on the formation of 1,2-cyclic forms of inositol phosphates^{19,20}, we identify this compound as cyclic InsP_3 . When this compound was incubated in the InsP_3 kinase assay, some ^{32}P radioactivity co-chromatographed on HPLC with an internal $^3\text{H}\text{-InsP}_4$ standard, showing that formation of ^{32}P -labelled 1,2-cyclic InsP_4 had occurred, though quenching the reaction with TCA (above) will have converted it to the non-cyclic form.

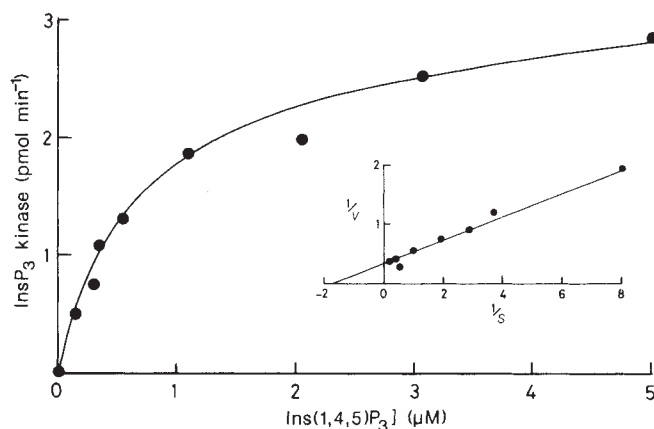


Fig. 3 Reaction velocity (V) of $\text{Ins}(1,4,5)\text{P}_3$ kinase plotted against substrate concentration (S). Assay conditions were as described in Fig. 2 legend, except that a 10-min incubation and a frozen brain supernatant were used. Each tube contained the equivalent of ~ 100 nl of brain cytosol. The data are from one of three identical experiments which gave similar results. Inset, a double-reciprocal plot of the same data.

The results show that the $\text{Ins}(1,4,5)\text{P}_3$ was phosphorylated in the 3 position only. When ^3H -*myo*- $\text{Ins}(1,4)\text{P}_2$ (Amersham) was incubated under the same conditions, we detected no InsP_3 or InsP_4 formation. Thus, we conclude that rat brain contains a specific $\text{Ins}(1,4,5)\text{P}_3$ 3-kinase.

Some basic characteristics of the enzyme were explored in brain homogenates or supernatants either fresh or stored frozen at -15°C ; the enzyme is very stable when frozen. These experiments were complicated by the presence of both membrane-bound and soluble enzymes which dephosphorylate $\text{Ins}(1,4,5)\text{P}_3$ (refs 10–12) and $\text{Ins}(1,3,4,5)\text{P}_4$ (see ref. 8). We were able to eliminate these to a large extent by including 3 mM 2,3-diphosphoglycerate in the incubations¹⁰. By using diluted broken cell suspensions and short incubation times, linear kinetics were achieved, though only with a pure enzyme will exact numbers for kinetic parameters be obtained.

The enzyme is ATP- and Mg^{2+} -dependent, and is predominantly soluble. Some activity remained in washed mem-

branes, but whether or not this represents entrapped activity is a difficult question to answer in brain homogenates¹³. We were unable to detect significant activity in human red cell membranes stored at -15°C which still possessed $\text{Ins}(1,4,5)\text{P}_3$ phosphatase activity¹⁰. Dialysis of soluble or particulate brain fractions against 0.15 M sucrose did not decrease the activity, and no effect of addition of GTP or a non-hydrolysable analogue (5'-guanylylimidodiphosphate) could be detected in supernatants or washed pellets at high or low substrate concentrations. The K_m of the kinase for $\text{Ins}(1,4,5)\text{P}_3$ was $\sim 0.6 \mu\text{M}$ and the $V_{\max}$ was 30 nmol per min per g fresh weight of tissue (Fig. 3). Thus the affinity of the kinase for $\text{Ins}(1,4,5)\text{P}_3$ is considerably higher than that of the 5-phosphatase^{10,12,14}. Moreover, it is probable that $\text{Ins}(1,4,5)\text{P}_3$ can reach a concentration in stimulated calls of the order of $1 \mu\text{M}^{4,7,15}$ ($= 1 \text{ nmol ml}^{-1}$), therefore a $V_{\max}$ of $\sim 0.5 \text{ nmol per s per ml}$ of cytosol readily accounts for the rapid formation of $\text{Ins}(1,3,4,5)\text{P}_4$ in stimulated brain slices⁸ and salivary glands⁹. This $V_{\max}$ is of a similar order to that of total InsP_3 phosphatase calculated by Storey *et al.*¹¹ in hepatocytes, suggesting that the phosphorylation and dephosphorylation of InsP_3 occur at similar rates (see also ref. 6).

As this novel pathway may help to curtail the second-messenger activity of $\text{Ins}(1,4,5)\text{P}_3$, and may also generate additional second messengers, it is important to know whether there is any hormonal control of the $\text{Ins}(1,4,5)\text{P}_3$ 3-kinase. As a first step we have examined the effects of pH and Ca^{2+} on the enzyme activity. The kinase was inactive at $\text{pH} \leq 6.5$, and above that pH the activity rose steeply to a broad optimum at ~ 8 ; this very steep rise in activity above pH 6.5 suggests the possibility of a pH regulation of activity, though within the pH change caused by, for example, growth factor-induced alkalization¹⁶, we calculate that the $\text{Ins}(1,4,5)\text{P}_3$ 3-kinase would be stimulated by only $\sim 20\%$ at the most. When the effect of calcium was explored in buffers at physiological ionic strength, pH and Mg^{2+} with ATP at $2 \text{ mM}^{17,18}$, kinase activity declined 2- to 3-fold over the range 10^{-7} – 10^{-5} M , suggesting that changes in kinase activity over physiological calcium ranges are also likely to be small.

We also examined the ability of the enzyme to phosphorylate the 1,2-cyclic form of $\text{Ins}(1,4,5)\text{P}_3$ (refs 19–21) (prepared from human blood cells; see Fig. 2 legend), and found that it would readily do so. The ^{32}P -labelled cyclic $\text{Ins}(1,4,5)\text{P}_3$ was of too low a specific activity to use at $<100 \mu\text{M}$, so we could not compare it in detail quantitatively with $\text{Ins}(1,4,5)\text{P}_3$ as a substrate; at $130 \mu\text{M}$ the rate of phosphorylation of $\text{Ins}(1,2\text{-cyclic-}4,5)\text{P}_3$ was 80% that of ^{32}P - $\text{Ins}(1,4,5)\text{P}_3$ prepared from the same red cells as the cyclic form. We consider it unlikely that the $\text{Ins}(1,2\text{-cyclic-}4,5)\text{P}_3$ was converted first to $\text{Ins}(1,4,5)\text{P}_3$ as studies on blood platelets and kidney indicate that the cyclic diester bond is not cleaved until the 4- and 5-phosphates have been removed¹⁴. Therefore, we suggest that the pathway outlined in Fig. 4 also holds if the 1-phosphate is cyclized to the 2-position of the inositol ring, and this in turn leads to the possibility that the differences observed between the effects of $\text{Ins}(1,4,5)\text{P}_3$ and those of its cyclic form on microinjection into photoreceptors²¹ reflect differences in their relative rates of phosphorylation (above) as opposed to dephosphorylation¹⁴.

Previous studies have characterized an inositol trisphosphate 5-phosphatase which terminates the second-messenger action of $\text{Ins}(1,4,5)\text{P}_3$ by converting it to $\text{Ins}(1,4)\text{P}_2$ (refs 10–12). In addition to this dephosphorylation pathway (Fig. 4), we have here identified a kinase which converts $\text{Ins}(1,4,5)\text{P}_3$ to $\text{Ins}(1,3,4,5)\text{P}_4$. The latter is then dephosphorylated to $\text{Ins}(1,3,4)\text{P}_3$ (ref. 8); the dephosphorylation of $\text{Ins}(1,3,4)\text{P}_3$ has not yet been studied, and its dephosphorylation to $\text{Ins}(1,4)\text{P}_2$ is possible but has not been established. As the two inositol trisphosphates accumulate in the parotid at comparable rates⁶, it is clear that a sizeable proportion of the $\text{Ins}(1,4,5)\text{P}_3$ released from $\text{PtdIns}(4,5)\text{P}_2$ is metabolized via this novel pathway to form both $\text{Ins}(1,3,4,5)\text{P}_4$ and $\text{Ins}(1,3,4)\text{P}_3$. The possibility remains that one or both of these compounds (or their 1,2-cyclic

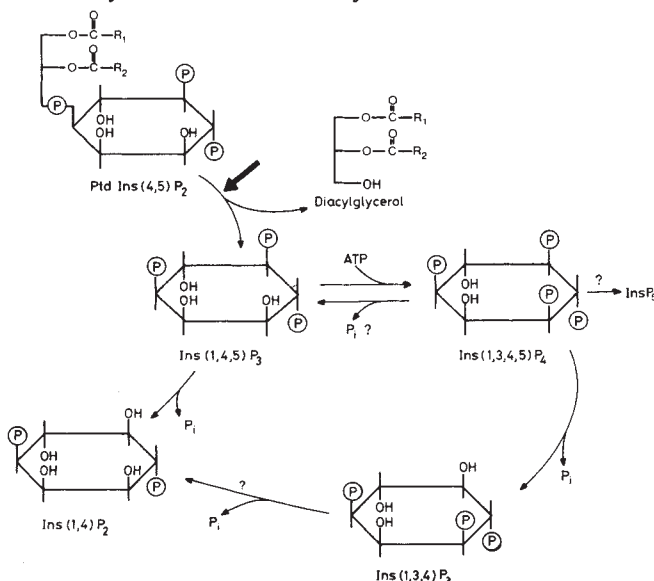


Fig. 4 The inositol tris/tetrakisphosphate pathway. The scheme outlines the suggested alternative routes of $\text{Ins}(1,4,5)\text{P}_3$ metabolism following its formation from $\text{PtdIns}(4,5)\text{P}_2$ by $\text{PtdIns}(4,5)\text{P}_2$ phosphodiesterase (large arrow). The scheme is based on the results of the present paper and on those of refs 5, 6 and 8, and also applies to the 1,2-cyclic forms of these inositol phosphates (see text for further discussion).

forms), which are unique to this pathway, may have second-messenger functions^{5,8}.

We thank Dr T. J. Rink of SK & F for the loan of HPLC apparatus.

Received 26 November 1985; accepted 26 February 1986.

- Berridge, M. J. *Biochem. J.* **220**, 345-360 (1984).
- Downes, C. P. & Michell, R. H. in *Molecular Mechanisms of Transmembrane Signalling* (eds Cohen, P. & Houslay, M. D.) 3-56 Elsevier, Amsterdam, 1985).
- Nishizuka, Y. *Nature* **308**, 693-698 (1984).
- Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
- Irvine, R. F., Letcher, A. J., Lander, D. J. & Downes, C. P. *Biochem. J.* **223**, 237-243 (1984).
- Irvine, R. F., Ånggård, E. E., Letcher, A. J. & Downes, C. P. *Biochem. J.* **229**, 505-511 (1985).
- Burgess, G. M., McKinney, J. S., Irvine, R. F. & Putney, J. W. *Biochem. J.* **232**, 237-243 (1985).
- Batty, I. R., Nahorski, S. R. & Irvine, R. F. *Biochem. J.* **232**, 211-215 (1985).
- Heslop, J. P., Irvine, R. F., Tashjian, A. H. & Berridge, M. J. *J. exp. Biol.* **119**, 395-402 (1985).
- Downes, C. P., Mussatt, M. C. & Michell, R. H. *Biochem. J.* **203**, 169-177 (1982).
- Storey, D. J., Shears, S. B., Kirk, C. J. & Michell, R. H. *Nature* **312**, 374-376 (1984).
- Connolly, T. M., Bross, T. E. & Majerus, P. W. *J. biol. Chem.* **260**, 7868-7874 (1985).
- Irvine, R. F. & Dawson, R. M. C. *J. Neurochem.* **31**, 1427-1434 (1978).
- Connolly, T. M., Wilson, D. B., Bross, T. E. & Majerus, P. W. *J. biol. Chem.* **261**, 122-126 (1986).
- Rittenhouse, S. E. & Sasson, J. P. *J. biol. Chem.* **260**, 8657-8660 (1985).
- Mills, G. B., Cragoe, E. J., Gelfand, E. W. & Grinstein, S. *J. biol. Chem.* **260**, 12500-12507 (1985).
- Downes, C. P. & Michell, R. H. *Biochem. J.* **202**, 53-58 (1982).
- Irvine, R. F., Letcher, A. J. & Dawson, R. M. C. *Biochem. J.* **218**, 177-185 (1984).
- Wilson, D. W., Bross, T. E., Sherman, W. R., Berger, R. A. & Majerus, P. W. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4013-4017 (1985).
- Dawson, R. M. C., Freinkel, N., Jungwala, F. B. & Clarke, N. *Biochem. J.* **122**, 605-607 (1971).
- Wilson, D. B. *et al. J. biol. Chem.* **260**, 13496-13501 (1985).
- Irvine, R. F., Hemington, N. & Dawson, R. M. C. *Biochem. J.* **164**, 177-180 (1977).

The signal sequence of nascent preprolactin interacts with the 54K polypeptide of the signal recognition particle

T. V. Kurzchalia*, M. Wiedmann*, A. S. Girshovich†, E. S. Bochkareva†, H. Bielka* & T. A. Rapoport*‡

* Central Institute of Molecular Biology, Academy of Sciences of GDR, 1115 Berlin-Buch, Robert-Roessle Str. 10, DDR

† Institute of Protein Research, Academy of Sciences of USSR, 142 292 Poustchino, Moscow region, USSR

Hydrophobic signal sequences direct the translocation of nascent secretory proteins and many membrane proteins across the membrane of the endoplasmic reticulum^{1,2}. Initiation of this process involves the signal recognition particle (SRP)^{3,4}, which consists of six polypeptide chains and a 7S RNA⁴ and interacts with ribosomes carrying nascent secretory polypeptide chains. In the case of amino-terminal, cleavable signal sequences, in the absence of microsomal membranes it exerts a site-specific translational arrest *in vitro*^{3,5,6}. The size of the arrested fragment (60-70 amino-acid residues) suggests that elongation stops when the signal sequence has emerged fully from the ribosome. However, a direct interaction between the signal sequence and SRP has not previously been demonstrated and has even been questioned recently⁷. We now show for the first time a direct interaction between the signal sequence of a secretory protein and a component of SRP, the 54K polypeptide (relative molecular mass (M_r) 54,000). This was achieved by means of a new method of affinity labelling which involves the translational incorporation of an amino acid, carrying a photoreactive group, into nascent polypeptides.

Our new method of affinity labelling allows us to identify the binding partners of a nascent polypeptide. A modified amino-acid residue, carrying a photoreactive group in its side chain, is incorporated into a protein, and crosslinking to the interacting component is induced by irradiation. The high selectivity of the aminoacyl tRNA synthetases, which normally only accept the

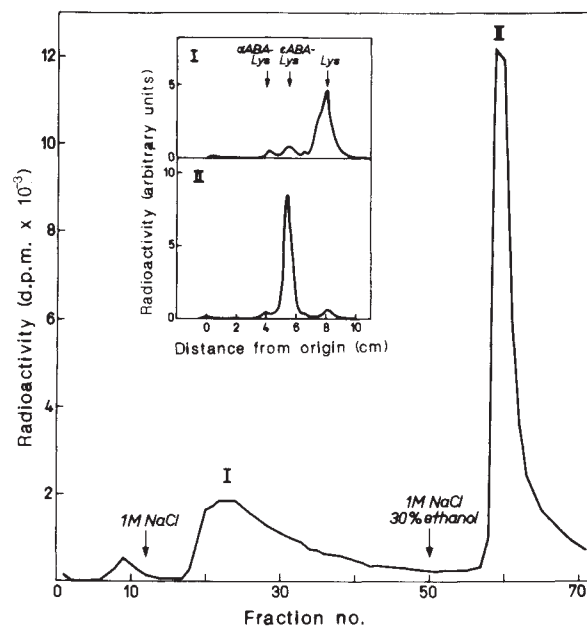


Fig. 1 Synthesis and purification of N^ϵ -*p*-azidobenzoyl-lysyl-tRNA (ϵ -ABA-Lys-tRNA). Unfractionated tRNA from baker's yeast was charged with ^{14}C -lysine by a crude gel-filtrated extract from rat liver¹⁹, and purified by phenol extraction and gel filtration. The specific radioactivity was about 60 pmol ^{14}C -lysine per A_{260} unit. All subsequent manipulations were carried out under dim light. For modification of the ϵ -position of lysine⁸, 300 A_{260} units of tRNA (18 nmoles Lys) were dissolved in 1 ml ice-cold water, and 1 ml of 0.2 M KH_2PO_4 , pH 7.0, and 1 ml 12 mM *N*-hydroxysuccinimide ester of *p*-azidobenzoic acid (synthesized as described in ref. 20) dissolved in tetrahydrofuran were added. The apparent pH was raised to 10.4 with 1 ml 0.12 M KOH. After 3 min, the tRNA was precipitated with ethanol and dissolved in a buffer containing 50 mM potassium acetate pH 5.0, 0.4 M NaCl and 10 mM magnesium acetate. The material was applied to a BD-cellulose column (1×15 cm) equilibrated with the same buffer. Elution was carried out as indicated in the figure. The fractions corresponding to peaks I (Lys-tRNA) and II (ϵ -ABA-Lys-tRNA) were precipitated with ethanol and aliquots were hydrolysed with 50 mM triethylamine for 1.5 h at 37°C. The samples were then subjected to thin layer electrophoresis at 300 V on cellulose sheets at pH 1.9 for 1 h (inset). ϵ -ABA- and α -ABA-lysine as well as the unmodified amino acid served as markers. The chromatograms were analysed using a radioactivity scanner.

20 naturally occurring amino acids, is overcome by introducing the modification to the amino acid after the tRNA has been charged. Johnson *et al.*⁸ have shown that acetylated lysyl-tRNA is accepted by the protein-synthesizing machinery.

Bovine preprolactin was chosen as a model to demonstrate an interaction between SRP and a signal sequence because it contains two lysine residues (positions 4 and 9) in front of the hydrophobic core of the signal peptide (positions 13-25)⁹. N^ϵ -*p*-azidobenzoyl-lysyl-tRNA (ϵ -ABA-Lys-tRNA) was synthesized and purified as described in Fig. 1 legend. The identity of the material was confirmed after hydrolysis by thin-layer electrophoresis on cellulose sheets (see Fig. 1 inset). The typical yield of the reaction was approximately 60%, and the final material consisted of more than 90% ϵ -ABA-Lys-tRNA.

To test whether modified ^3H -Lys-tRNA is accepted by the protein-synthesizing machinery, the tRNA was added to a cell-free translation system from wheat germ programmed with globin or pituitary messenger RNA. In both cases incorporation of radioactive lysine in trichloroacetic acid (TCA)-precipitable material was observed. Moreover, the kinetics of incorporation were similar for unmodified and ϵ -modified Lys-tRNA (Fig. 2a). SDS-polyacrylamide gel electrophoresis (SDS-PAGE) demonstrated that the radioactivity was present in distinct trans-

‡ To whom correspondence should be addressed.

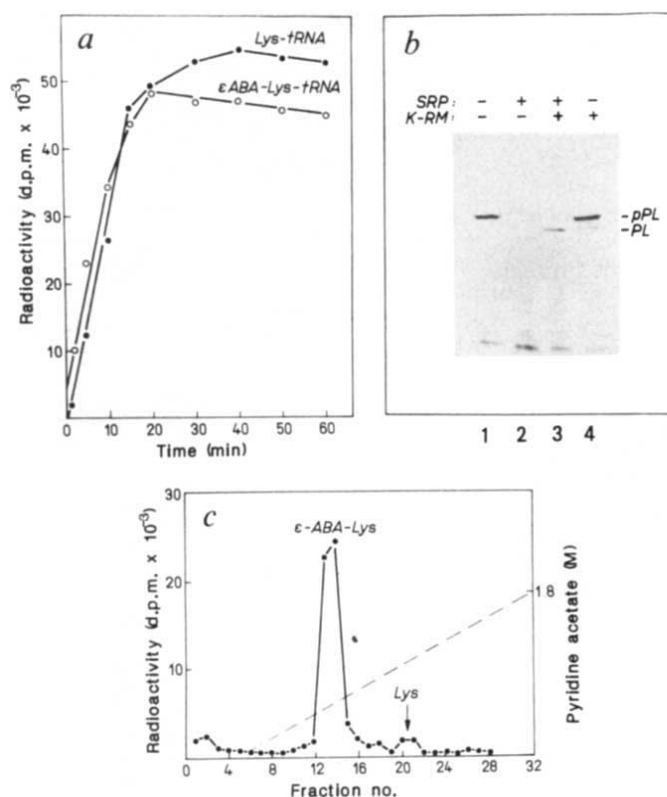


Fig. 2 Translational incorporation of ϵ -ABA- 3 H-Lys into newly synthesized polypeptides. *a*, Kinetics of incorporation of 3 H-lysine and ϵ -ABA- 3 H-Lys into TCA-precipitable material. *b*, SDS-PAGE of cell-free synthesized translation products directed by pituitary mRNA. *c*, Ion-exchange chromatography of hydrolysed preprolactin synthesized in the presence of ϵ -ABA- 3 H-Lys-tRNA. pPL, preprolactin; PL, prolactin; K-RM, high-salt-washed microsomes. **Methods.** ϵ -ABA- 3 H-Lys-tRNA was obtained as described in Fig. 1 legend except that 3 H-lysine (80 Ci mmol $^{-1}$) was used. Cell-free translation of pituitary RNA, isolated from lactating cows, was carried out in a wheat germ system as described previously^{14,21} except that dithiothreitol was replaced by 2 mM 2-mercaptoethanol to prevent decomposition of the azido group. About 1.5×10^5 d.p.m. of ϵ -ABA- 3 H-Lys-tRNA (0.8 pmoles) was added to 50 μ l of a translation mixture. After incubation for the indicated times, the samples were precipitated by TCA, heated to 95 $^{\circ}$ C for 10 min and the radioactivity was determined (*a*). SDS-PAGE in a 10–20% linear acrylamide gel and fluorography were performed as described elsewhere^{14,21}. SRP (30 units) (lanes 2 and 3) and high-salt-washed microsomes (2 equiv.) (lanes 3 and 4) were added to the translation mixture. SRP was isolated from dog pancreas and purified by hydrophobic chromatography and sucrose density centrifugation²². Exhaustive hydrolysis of preprolactin synthesized in the presence of ϵ -ABA- 3 H-Lys-tRNA was performed by combined action of trypsin, chymotrypsin and aminopeptidase M for a total time of 48 h as described by Johnson *et al.*⁸ for globin. The pH was then adjusted to 2 and 100 μ l of hydrolysate were applied to a column (0.4 $\times$ 13 cm) filled with spherical ion-exchange resin type M-82 (Beckman). Elution was done with a linear gradient of 50–2,000 mM pyridine/acetic acid at 45 ml h $^{-1}$ and 15 atm. Fractions of 1.2 ml were collected and the radioactivity in 1 ml counted. The column was calibrated with ϵ -ABA- 3 H-Lys and 3 H-lysine.

lation products (see Fig. 2*b*), (results for globin not shown). When the synthesized preprolactin was hydrolysed enzymatically, more than 95% of the radioactivity was recovered in ϵ -ABA-Lys (Fig. 2*c*). These data show that a modification of lysine, which increases the estimated size of the side chain from 10 Å to 16 Å, does not impair its acceptability by the translational apparatus or its ability to pass through the ribosome.

As a precondition for further experiments, it was necessary to show that the incorporation of modified lysine residues into a secretory protein does not interfere with the activity of SRP.

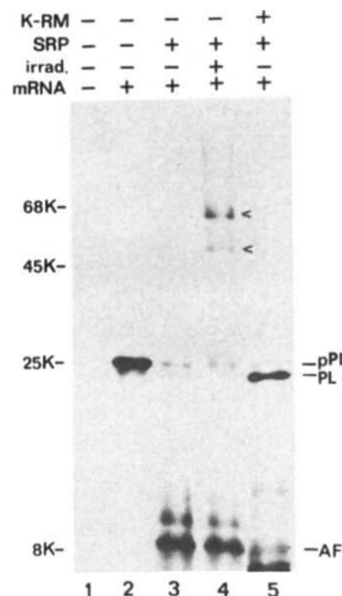


Fig. 3 Crosslinking of components contained in the arrested complex. Cell-free translation of pituitary RNA was synchronized by addition of 6 mM 7 mGp 1 min after the start of translation³. The translation mixture contained about 16 pmol of ϵ -ABA-Lys-tRNA and 88 μ Ci of 3 S-methionine as label per 50 μ l. SRP (30 units) and K-RM (2 equiv.) were added as indicated. Translation was carried out for 15 min, 2 μ l of 1 mM methionine was added and the incubation continued for 5 min. The mixture was then passed through a Sepharose 6B-CL column (1 $\times$ 20 cm) equilibrated with a buffer containing 50 mM HEPES pH 7.5, 150 mM potassium acetate, 5 mM magnesium acetate and 0.01% Nikkol. Fractions corresponding to the void volume were collected and 200- μ l aliquots were irradiated with light of 313 nm wavelength for 2 min at a dose of 10^5 erg mm $^{-2}$. The samples, corresponding to a volume of 50 μ l of original translation mixture, were precipitated by TCA before electrophoresis in a 10–20% SDS gel. An equal amount of non-irradiated material served as control. The samples applied to lanes 1, 2 and 5 (12.5 μ l translation assays) were directly processed for SDS-PAGE. *M_r* standards in the SDS-PAGE were bovine serum albumin, ovalbumin and chymotrypsin. AF, arrested fragment (8K); pPL, preprolactin; PL, prolactin.

Figure 2*b* demonstrates that SRP function is indeed retained³: in the absence of microsomal membranes (lane 2), an inhibition of preprolactin translation was observed by SRP which was released by the addition of high-salt-washed microsomes (lane 3). The latter were inactive by themselves, as they lacked SRP (Fig. 2*b*, lane 4). These results correspond to those with unmodified preprolactin (not shown).

For crosslinking experiments, a synchronized translation was carried out in the presence of 3 S-methionine and ϵ -ABA-Lys-tRNA, the 3 S label being used to facilitate product analysis. SRP was added to produce elongation arrest. In these conditions the formation of an arrested complex occurs with SRP bound to ribosomes carrying nascent polypeptide fragments of about 6,000–8,000 *M_r* (see Fig. 3, lane 3)^{3,5}. After incubation, the mixture was passed through a Sepharose 6B-CL column which separates the arrested complex from unincorporated tRNA and completed preprolactin. Irradiation of the arrested complex resulted in the appearance in a SDS-gel of a main band of ~60K *M_r* and a minor and variable band of 50K which was not studied further (Fig. 3, lane 4). The radioactivity in the two bands formed about 5% of the total exposed to irradiation. Neither band was seen if ϵ -ABA-Lys-tRNA or SRP was omitted from the assay, nor did they appear if SRP was added after translation (data not shown). Because the main radioactive component before irradiation is the arrested fragment (AF) (see lane 3), the high-*M_r* bands should represent products crosslinked with it. Consequently, the interacting species may be assumed to have *M_r*s

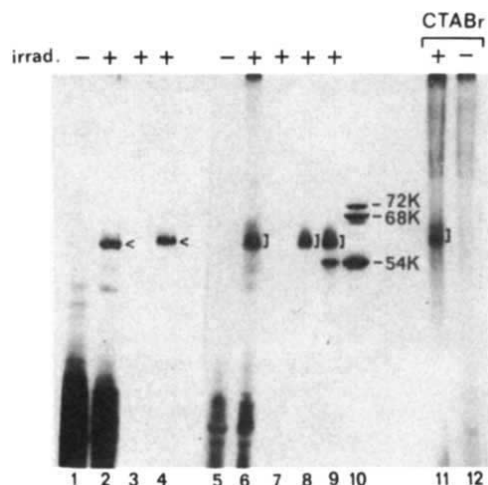


Fig. 4 Crosslinking of the arrested nascent polypeptide fragment to the 54K protein of SRP. Synchronized cell-free translation was carried out in the presence of SRP (15 U per 25 μ l). Initiation of translation was stopped by addition of 7 mGp either after 1 min (lanes 1–4) or, in a different experiment, after 2 min (lanes 5–9 and 11–14). The samples (except those applied to lanes 11 and 12) were passed through a Sepharose column as described in Fig. 3 legend and portions were irradiated. Some samples were processed directly for SDS-PAGE (lanes 1, 2, 5, 6) and to others antibodies were added. For immunoprecipitation, the volume of the translations was increased 6-fold (lanes 3, 4) and 10-fold (lanes 7–9), respectively. The antigen–antibody complexes were collected with protein A–Sepharose, dissociated with SDS and 2-mercaptoethanol and applied to SDS-PAGE (10–15%). Affinity-purified IgG was isolated from a rabbit antiserum directed against the 54K polypeptide of SRP¹⁰ and added to the test samples (4 μ g for sample 4 and 8 μ g for samples 8 and 9). To test for the specificity of immunoprecipitation of the 54K polypeptide, sample 9 also received about 10^4 d.p.m. of iodinated SRP (lane 10) prepared as described in ref. 5. Non-immune IgG was added to all samples (200 μ g for samples 3 and 4 and 400 μ g for 7–9). Lanes 11 and 12 shows the material precipitated directly by cetyltrimethylammonium bromide (CTABr)^{11,12} after translation of an irradiated and non-irradiated sample, respectively.

of 52–54K (for the 60K band) and 42–44K (for the 50K band), respectively.

To investigate whether the main band represented the cross-linked product of the radioactive arrested fragment with the 54K polypeptide of SRP, immunoprecipitation experiments were carried out using affinity-purified anti-54K protein antibodies (Fig. 4)¹⁰. As Fig. 4 shows, the 60K crosslinking product is indeed recognized by the antibodies (lanes 4, 8, 9). Controls with non-immune serum confirmed the specificity of the immunoprecipitation (Fig. 4, lanes 3, 7). In some experiments there was a broadening of the 60K band in the direction of higher M_r (up to 66K, compare lane 4 with lanes 8 or 9 in Fig. 4); apparently, SRP can interact with ribosomes carrying nascent chains of 50–70 to about 110 amino-acid residues in length, and crosslinking occurs with the N-terminus, which is common to all sizes of peptide. Figure 4 also shows that the product is precipitated by cetyltrimethylammonium bromide¹¹ (compare lanes 11 and 12) which is further evidence that crosslinking occurred with a nascent polypeptide chain still bound to tRNA¹².

Our results show that the arrested fragment of preprolactin, the main translation product of pituitary mRNA, interacts with the 54K polypeptide of SRP. The only lysine residues contained in the shortest arrested fragment are those in positions 4 and 9 and these must therefore be responsible for crosslinking by the incorporated photoreactive group. We therefore conclude that it is the signal sequence which interacts with the 54K protein of SRP. Our results support previous suggestions^{3,13–15} that the

hydrophobic signal sequences initially interact with a special protein receptor contained in SRP, rather than with the phospholipid bilayer^{16–18}.

The technique described here may be of general use, for not only can photoreactive groups be translationally incorporated into proteins, but also other chemical groups may be used for their labelling or isolation. This method may serve to identify the environment of a protein during its passage through a membrane or to label receptors involved in protein transport into cell organelles (work in progress).

We thank P. Walter for the generous gift of antibodies to the 54K protein of SRP, R. Dettmer for help with ion-exchange chromatography of protein hydrolysates, and S. M. Rapoport and H. Welfle for critical reading of the manuscript.

Received 19 November; accepted 16 December 1985.

1. Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1496–1500 (1980).
2. Rapoport, T. A. & Wiedmann, M. *Curr. Topics Membrane Transport* **24**, 1–63 (1985).
3. Walter, P. & Blobel, G. *J. Cell Biol.* **91**, 557–561 (1981).
4. Walter, P. & Blobel, G. *Nature* **299**, 691–698 (1982).
5. Walter, P., Ibrahim, I. & Blobel, G. *J. Cell Biol.* **91**, 545–550 (1981).
6. Meyer, D. I., Krause, E. & Dobberstein, B. *Nature* **297**, 647–650 (1982).
7. von Heijne, G. *FEBS Lett.* **190**, 1–5 (1985).
8. Johnson, A. E., Woodward, W. R., Herbert, E. & Menninger, J. R. *Biochemistry* **15**, 569–575 (1976).
9. Sasavage, N. L., Nilson, J. H., Horowitz, S. & Rottman, F. M. *J. biol. Chem.* **257**, 678–681 (1982).
10. Walter, P. & Blobel, G. *J. Cell Biol.* **97**, 1693–1699 (1983).
11. Hobden, A. H. & Cundliffe, E. *Biochem. J.* **170**, 57–61 (1978).
12. Gilmore, R. & Blobel, G. *Cell* (in the press).
13. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 852–862 (1975).
14. Prehn, S., Tsamaloukas, A. & Rapoport, T. A. *Eur. J. Biochem.* **107**, 185–195 (1980).
15. Finkelstein, A. V., Bendzko, P. & Rapoport, T. A. *FEBS Lett.* **161**, 176–179 (1983).
16. von Heijne, G. & Blomberg, C. *Eur. J. Biochem.* **97**, 175–181 (1979).
17. Engelman, D. M. & Steitz, T. A. *Cell* **23**, 411–422 (1981).
18. Wickner, W. A. *Rev. Biochem.* **48**, 23–45 (1979).
19. Bommer, U. A., Henske, A. & Bielka, H. *Acta biol. med. germ.* **37**, 1363–1376 (1978).
20. Galaray, R. E., Craig, L. C., Jamieson, J. D. & Printz, M. P. *J. biol. Chem.* **249**, 3510–3518 (1974).
21. Wiedmann, M., Huth, A. & Rapoport, T. A. *Nature* **309**, 637–639 (1984).
22. Walter, P. & Blobel, G. *Meth. Enzym.* **96**, 682–691 (1983).

A consensus amino-acid sequence repeat in *Torpedo* and mammalian Ca^{2+} -dependent membrane-binding proteins

Michael J. Geisow*, Ulrich Fritschel†, J. Mark Hexham‡, Bret Dash* & Tolu Johnson*

* Laboratory of Protein Structure, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

† Abteilung Neurochemie, Max-Planck-Institut für Biophysikalische Chemie, Göttingen, FRG

‡ Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

A group of calcium-binding proteins which bind to biomembranes has recently been identified in widely different cells and tissues (refs 1–7, reviewed in ref. 8). Three of these proteins (p70, p36 and p32.5) cross-react with antiserum to calelectrin, a Ca^{2+} -binding protein (relative molecular mass 34,000 (34K)) from the ray *Torpedo marmorata*^{3,4}, giving rise to their designation as calelectrin-related proteins³. We now report that calelectrin, p36 and p32.5 contain a 17-amino-acid consensus sequence which is conserved and present in multiple copies. We suggest that this sequence may be common to other members of this new group of Ca^{2+} -binding proteins and may underlie their unusual mode of combination with biomembranes.

p70 and p32.5, found in liver and adrenal, have been characterized^{3,4} and p36, from adrenal, has been shown to correspond to the 36K major intracellular substrate for pp60^{c-src} tyrosine kinase^{9,4}. The native protein consists of a 36 dimer and two 10K subunits¹⁰. p36 has previously been recognized as a Ca^{2+} -

binding component of the brush border membranes of intestinal epithelial cells⁹. p70 and p32.5 are not related to calmodulin because they are heat-labile³ and fail to bind in purified form to phenothiazine affinity columns⁴ or to expose hydrophobic sites in response to Ca^{2+} (refs 11, 12). We have studied the membrane-binding properties of p32.5, which we have named endonexin by analogy with the Ca^{2+} -binding protein synexin and because of its endoplasmic-reticulum-like immunolocalization in fibroblasts⁴. Homogeneous liver endonexin bound reversibly to liposomes formed from equimolar amounts of phosphatidylcholine (PC) and phosphatidylinositol (PI), phosphatidylethanolamine (PE) (Fig. 1a) or phosphatidic acid (PA) (not illustrated) between 1 μM and 10 μM free Ca^{2+} . Magnesium ions could not replace calcium at any concentration and inhibited Ca^{2+} -dependent binding above 10 mM Mg^{2+} . Endonexin did not bind to liposomes containing either PC alone (Fig. 1a) or equimolar mixtures of PC with phosphatidylserine (PS), sphingomyelin or cholesterol (not shown). Binding to liposomes closely resembled the previously reported interaction with secretion vesicles (chromaffin granules) of adrenal medulla³. In particular, binding to granule membranes³ and liposomes (Fig. 1b) was strongly dependent on the presence of free Ca^{2+} .

At least five Ca^{2+} -binding proteins were immunostained on Western blots by polyclonal calelectrin antiserum⁴, and endonexin and p70 were recognized by monoclonal antibodies produced in response to challenge with pure calelectrin (U.F., unpublished results). We considered whether this surprising degree of cross-reactivity might be due to a hapten-like modification of proteins which were otherwise structurally distinct. Several proteins became bound to membranes following the post-translational addition of amino-terminal myristic acid. In particular, a proportion of p36 carries this modification¹³, and if myristic acid formed part of an epitope recognized by anti-calelectrin, this could explain the cross-reaction with p36 and possibly with the other proteins described.

We compared our adrenal Ca^{2+} -binding proteins with the B subunit of bovine calcineurin, pp60^{v-src} and pp60^{c-src}, all three of which are myristylated^{14,15}. None of these known myristic-acid-containing proteins was immunostained by anti-calelectrin on Western blots, although the pp60^{c-src} preparation contained a positively staining contaminant of 36,000 relative molecular mass that was possibly p36. Phosphotyrosine was also ruled out as a common hapten when 10 mM O-phospho-L-tyrosine was found to have no quantitative effect on the binding of calelectrin antibodies to adrenal Ca^{2+} -binding proteins on Western blots.

We previously noted that proteolytic digestion of each adrenal Ca^{2+} -binding protein gave a small number of peptides which reacted with calelectrin antiserum⁴, indicating that a common determinant might be found by partial sequence analysis. Pure

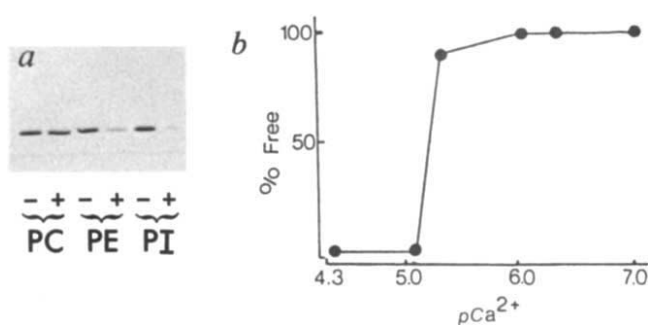
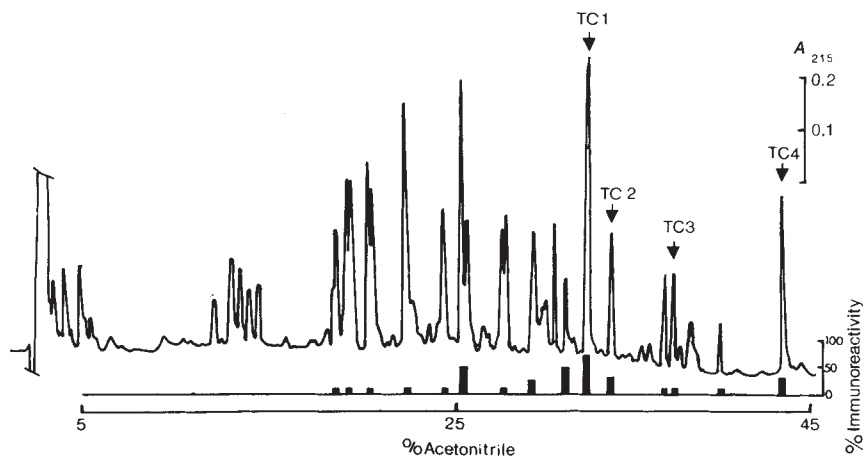


Fig. 1 Binding of endonexin to liposomes. *a*, Lipid dependence. Supernatants at 0.1 μM Ca^{2+} (-) and 10 μM Ca^{2+} (+) were run on an SDS-polyacrylamide gel. In this experiment the lipids were PC (pure component), PE (1:1 PC:PE) and PI (1:1 PC:PI). *b*, Ca^{2+} dependence. Equal aliquots of supernatants were obtained as described below where the liposomes were 1:1 PC:PI. The ordinate (% Free) indicates the percentage of endonexin remaining in the supernatant at the calcium concentration ($p\text{Ca}$) indicated. **Methods.** Liposomes were made by swelling dried lipids in 100 mM KCl/5 mM EDTA/5 mM MgCl_2 /20 mM HEPES (pH 7.5) buffer. After sonication, $p\text{Ca}$ was adjusted with CaCl_2 and checked by a Radiometer Ca^{2+} electrode. Aliquots (5 μg) of endonexin were mixed with 200 μg lipid and, after 15 min at 20 $^\circ\text{C}$, centrifuged at 100,000 g_{av} in an airfuge.

calelectrin, endonexin and p36 (the 36K subunit of p36) were reduced, alkylated and digested with cyanogen bromide or trypsin, and CNBr fragments of endonexin were resolved by HPLC. A single CNBr fragment of endonexin (CNBr6) was immunostained by anti-calelectrin in dot-blot assays and yielded the partial amino-acid sequence shown in Fig. 3. The presence of an internal methionine at position 14 and the low initial yield of this peptide (150 pmol) indicate that it arose from incomplete cleavage. We do not know the length of the fragment, but its composition and elution position suggest that it is mainly complete (see Fig. 3).

Tryptic digests of calelectrin and endonexin were limiting and gave reproducible fragmentation patterns by HPLC. Figure 2 shows the separation of calelectrin tryptic peptides. Equal proportions of each peak were tested for immunoreactivity with calelectrin antiserum by enzyme-linked immunosorbent assay (ELISA). The peptides TC1-4, together with random tryptic peptides of endonexin and CNBr peptides of p36, were sequenced to completion by automated Edman degradation. A homologous sequence repeat was found in each protein (Fig. 3). The ELISA results (Fig. 2) clearly demonstrate that calelectrin antiserum contains immunoglobulins which recognize the repeat

Fig. 2 Isolation of calelectrin tryptic peptides and their immunoreactivity with anti-calelectrin. Reduced, S-carboxymethylated calelectrin (5 nmol) was digested for 48 h with trypsin (1:50 w/w protein) and peptides were isolated by chromatography on a Waters μ Bondapak C₁₈ reverse-phase column. Equal proportions of pooled peaks were dried on ELISA plates and their immunoreactivity with calelectrin antiserum was expressed as a percentage of that of 0.1 μg of undigested protein.



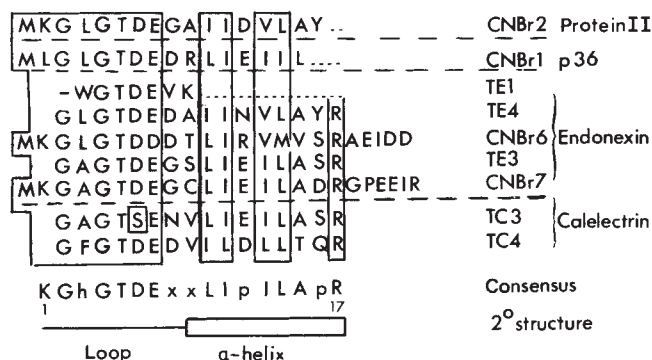


Fig. 3 Amino-acid sequences of endonexin CNBr peptides and tryptic peptides from endonexin (TE) and calelectrin (TC) were sequenced to completion using an Applied Biosystems model 470A gas-phase protein sequencer. Endonexin (p32.5) was purified from bovine liver⁴, p36 from pig mesenteric lymph nodes²¹ and protein II from pig intestinal epithelia¹⁰. Protein II cross-reacts immunologically with endonexin²². Methionines are shown at the amino termini of all CNBr peptides, because the protein amino termini appear to be blocked. Similarly, the tryptic peptides must each be preceded either by lysine or arginine. In the consensus sequence, h = hydrophobic residue, p = polar residue and x = variable residue. Secondary structure was predicted by a modification of the method of Chou and Fasman (Roberts and M.J.G., unpublished).

sequences TC3 and TC4 (Fig. 3). The amino-acid sequence of the most immunoreactive peptide, TC1, is unrelated to any sequence determined for mammalian proteins (not shown). The tryptic peptides and CNBr fragments sequenced so far account for approximately half of endonexin (150 residues) and more than 50% of these occur in the conserved repeat. It seems likely that more repeats will emerge after the complete sequence has been determined, since internal basic amino acids have led to the fragmentation of the putative repeat in TE1, the repeat contained within CNBr6 and in other sequences not shown. Substantially less calelectrin sequence has been determined (63 residues) and more repeats could be present.

The consensus repeat sequence reveals a pattern of conserved glycines, polar, nonpolar and charged residues within an exact length of 17 amino acids (Fig. 3). There is no statistically significant difference between the repeats in calelectrin and endonexin, which together have a homology of 68%. A search of the current protein sequence databases (3,500 sequences) failed to find significant homology between the sequences in Fig. 3 and any known protein.¹⁶ No other significant homologies have been found and so the similarity of the primary structure of the proteins is currently restricted to the repeated sequences shown in Fig. 3.

When terbium was used as a spectroscopic probe for Ca^{2+} -binding sites in endonexin and calelectrin, multiple sites were present in both proteins^{11,12}. Fluorescence emission spectra from protein-bound Tb^{3+} gave an excitation maximum at 285 nm, characteristic of tryptophan residues close to Tb^{3+} sites^{11,12}. As there are at least five conserved repeats in endonexin, one of which (TE1) contains tryptophan, these repeats may form part of Ca^{2+} -binding sequences, despite lack of homology with the E-F hand of many Ca^{2+} -binding proteins¹⁶. Multiple binding sites are also indicated by the cooperative binding to liposomes.

It is striking that endonexin combines only with liposomes containing lipids possessing marked asymmetry in the bilayers of cell membranes. PE, PI and PA are found predominantly in the cytoplasmic leaflets of plasma or endoplasmic reticulum membranes¹⁷. PC, which is not bound by endonexin, has a less marked asymmetrical distribution. PS, which is almost exclusively present in the cytoplasmic leaflet of the plasma membrane, is not bound by endonexin, but is bound by the analogous Ca^{2+} -

binding protein synexin, which seems to be co-distributed with endonexin and other proteins described here^{5,18}.

The sarcoma virus tyrosine kinase substrate, p36, has also been shown to bind to liposomes containing PS or PI¹⁹. The presence of lipid shifts the Ca^{2+} -dependence of tyrosine phosphorylation from the millimolar to the micromolar level¹⁹. The demonstration of Ca^{2+} -dependent liposome binding by p36, endonexin and calelectrin²⁰, together with the conserved sequences described here, strengthens previous suggestions that these proteins are related in their structure and functions^{3,4,8}. We speculate that the common sequence represents part of either a new type of Ca^{2+} -binding site or a lipid-binding site and that this feature will be common to other members of this new family of Ca^{2+} -binding proteins. The established lipid¹⁹ and cytoskeleton¹⁰ binding properties of p36 suggests that this protein and possibly the others described, could cross-link lipids in the bilayer to membrane cytoskeletal elements.

We acknowledge the support of EEC 'Stimulation Action' grant STI-U69-S-UK from the directorate general for Science Research and Development (for U.F.). We thank Drs V. Witzemann and V. P. Whittaker for critical comment on the manuscript, Drs P. Cohen (Dundee) and A. I. Magee (National Institute for Medical Research) for gifts of calcineurin B and pp60, respectively, N. Totty (Imperial Cancer Research Fund) for some sequence determinations, and A. Harris and J. Childs for technical assistance.

Note added in proof: After submission of this paper, the protein sequence deduced from human lipocortin cDNA was published²³. This phospholipase A_2 inhibitor contains the consensus sequence described.

Received 16 December 1985; accepted 26 February 1986.

- Walker, J. H. *J. Neurochem.* **39**, 815-823 (1982).
- Walker, J. H., Obrocki, J. & Südhof, T. C. *J. Neurochem.* **41**, 139-145 (1983).
- Südhof, T. C., Ebbecke, M., Walker, J. H., Fritsche, U. & Boustead, C. *Biochemistry* **23**, 1103-1109 (1984).
- Geisow, M. J. *et al.* *EMBO J.* **3**, 2969-2974 (1984).
- Creutz, C. E. *et al.* *J. Biol. Chem.* **258**, 14664-14674 (1985).
- Owens, R. J., Gallagher, C. J. & Crumpton, M. J. *EMBO J.* **1**, 1167-1170 (1984).
- Moore, P. B. & Dedman, J. R. *J. Biol. Chem.* **257**, 9663-9667 (1982).
- Owens, R. J. & Crumpton, M. J. *Biochem. Essays* **1**, 61-63 (1984).
- Gerke, V. & Weber, K. *EMBO J.* **3**, 227-233 (1984).
- Gerke, V. & Weber, K. *J. Biol. Chem.* **260**, 1688-1695 (1985).
- Südhof, T. C., Walker, J. H. & Fritsche, U. *J. Neurochem.* **44**, 1302-1307 (1985).
- Südhof, T. C. *Biochem. biophys. Res. Commun.* **123**, 100-107 (1984).
- Soric, J. & Gordon, J. A. *Science* **230**, 563-566 (1985).
- Sefton, B. M., Trowbridge, I. S., Cooper, J. A. & Scolnick, E. M. *Cell* **31**, 456-474 (1982).
- Aitken, A. *et al.* *FEBS Lett.* **150**, 314-318 (1982).
- Tufty, R. M. & Kretsinger, R. H. *Science* **187**, 167-169 (1975).
- Op den Kamp, J. A. F. *A. Rev. Biochem.* **48**, 47-71 (1979).
- Morris, S. J., Hughes, J. M. X. & Whittaker, V. P. *J. Neurochem.* **39**, 529-536 (1982).
- Glenney, J. R. *FEBS Lett.* **192**, 79-82 (1985).
- Südhof, T. C., Walker, J. H. & Obrocki, J. *EMBO J.* **1**, 1167-1170 (1982).
- Hexham, J. M., Totty, N. F., Waterfield, M. D. & Crumpton, M. J. *Biochem. biophys. Res. Commun.* **134**, 248-254 (1986).
- Davies, A. A. & Crumpton, M. J. *Biochem. biophys. Res. Commun.* **128**, 571-577 (1985).
- Wallner, B. P. *et al.* *Nature* **320**, 77-81 (1986).

Erratum

Danish Basin subsidence by Triassic rifting on a lithospheric cooling background

K. Sørensen

Nature **319**, 660-663 (1986)

IN this letter Figs 2 and 3 were transposed; the legends are correct as printed.

Corrigendum

Electron microscopy of frozen-hydrated biological material

M. Stewart & G. Vigers

Nature **319**, 631-636 (1986)

IN the numbered references 11, 16, 28, 38, 41 and 47, the author's name should read McDowall, A. W. The correct form is seen in ref. 39.

Founder of phage genetics

Max Perutz

Licht und Leben: Ein Bericht über Max Delbrück, den Wegbereiter der Molekularbiologie.* By Peter Fischer.

Universitätsverlag Konstanz, Postfach 6632, 7750 Konstanz, FRG:1985. DM 34.80.

THE EARLY conceptual development of molecular biology was dominated by two physicists: Max Delbrück and Francis Crick. Both owe their fame to a small number of seminal papers and their influence to their formidable powers of imagination and argument. Peter Fischer's carefully researched biography explains Delbrück's scientific work in words and diagrams intelligible to the layman, and paints a vivid, thoughtful, affectionate, humorous and balanced portrait of the man. It is to be hoped that the book will soon appear in English, and thus be known to a wider audience.

I learnt much that I had not known, such as Delbrück's early association with Otto Hahn and Lise Meitner, who engaged him as a theoretician to help them interpret their bombardment of uranium with neutrons. Delbrück failed to grasp the meaning of their results, but he applied the target theory learnt from them to the Russian geneticist Timoféeff-Ressovsky's mutagenic quantum yield of X-rays and calculated that the gene must be a molecule containing no more than a few hundred atoms. Even though his estimate was wrong because he neglected the effects caused indirectly by the generation of free radicals in the surrounding medium, the paper secured him a Rockefeller Fellowship to go and study in Pasadena; it also stimulated Schrödinger to write his influential book *What is Life?*, published by Cambridge University Press in 1946, in which he predicted the gene to be a molecule with an aperiodic structure; and it made a young medical graduate learning physics in Enrico Fermi's laboratory in Rome decide to work with Delbrück on the nature of the gene. His name was Salvatore Luria.

Fischer recounts that Delbrück's entry into biology was inspired by a lecture, "Light and Life", by his teacher Niels Bohr. Bohr predicted that the study of life at the atomic level would lead to a paradox similar to that posed earlier by atomic spectra, a paradox that was resolved only by the new quantum mechanics:

The existence of life must be considered as an

elementary fact that cannot be explained, but must be taken as a starting point in biology, in a similar way as the quantum of action, which appears as an irrational element from the point of view of classical mechanical physics, taken together with the existence of elementary particles, forms the foundation of atomic physics. The asserted impossibility of a physical or chemical explanation of the function peculiar to

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Fruitful partnership — Salvador Luria (standing) and Max Delbrück at Cold Spring Harbor Laboratory in 1953.

life would be . . . analogous to the insufficiency of the mechanical analysis for the understanding of the stability of atoms [Niels Bohr, *Nature* 131, 458 (1933)].

Delbrück himself has described the search for this "Elementary Fact of Life" as the sole motive behind all of his work. He should have listened instead to Linus Pauling, with whom he actually published a joint paper in 1940, and who realized that the biological "quantum of action" is the hydrogen bond, which accounts for most biochemical reactions without having to invoke any new "Elementary Facts".

In 1937, for scientific rather than ideological or racial reasons, Delbrück left Berlin for T.H. Morgan's laboratory at the California Institute of Technology. He hoped to reduce the genetics of *Drosophila* to simple physical principles, but was disappointed to find no quantitative data susceptible to theoretical interpretation. He was about to give up when he discovered that in the basement of the same building another biologist, E.L. Ellis, was working on bacteriophages in *Escherichia*

coli. A glance at Ellis's plaques convinced Delbrück that the bacteriophage was the hydrogen atom of biology for which he had been looking, and that study of it might lead him to the "Great Paradox of Life". He and Ellis soon discovered that a single phage adsorbed to a single bacterium multiplies "upon or within" that bacterium until it bursts with the release of an average of 60 progeny phages; such a mechanism had been proposed by d'Herelle, but never proved, while others thought they had found evidence in favour of continuous release of phage by infected bacteria. By clear thinking and application of the simple theory of exponential growth, Delbrück and Ellis opened up the pathway to the analysis of phage genetics. Delbrück soon attracted an enthusiastic band of disciples, who formed the flourishing phage school that met each summer at Cold Spring Harbor in New York state.

In 1938, Luria obtained an Italian government fellowship to work with Delbrück in Pasadena, but Mussolini's racial laws annulled that, and it was not until September 1940 that Luria reached New York, now as a refugee. He sought out Delbrück, who had become an instructor at Vanderbilt University in Nashville, Tennessee. They first worked together on interference between two bacterial viruses acting on the same host, where Delbrück hoped to find something analogous to Pauli's exclusion principle in physics. The paper which was to earn them the Nobel Prize 26 years

later was conceived in 1942, after Luria had found a job at Bloomington, Indiana. Luria tried to discover whether bacterial resistance to phage infection was caused by an adaptive change, as many believed, or whether it arose from mutations. He was perplexed by the extreme variability of the numbers of resistant bacteria present in different cultures of the same organisms, until the correct explanation dawned upon him one night at a dance while watching a game machine. If the change from susceptibility to resistance was a random event due to mutations, then a mutation occurring early in the life of a culture would give rise to a large clone of resistant bacteria, while several mutations arising later would each produce only small clones. Luria wrote to Delbrück, telling him of his idea; Delbrück put it into mathematical form and proved rigorously that the distribution of resistant bacteria in Luria's different cultures was consistent only with their being due to random mutations, and that these mutations occurred with a constant frequency of 2.45×10^{-8} per bacterial div-

CHSL Library Archives. Reproduced from *The DNA Story* (W.H. Freeman, 1981).

**Light and Life: An Account of Max Delbrück, Pioneer of Molecular Biology.*

ision. These results opened up the field of bacterial genetics. Just like Delbrück and Ellis's earlier results on phage, they involved nothing that could not have been found out years earlier; the only new ingredient was clear thought.

In the ecology of science the opening of a new habitat immediately attracts a crowd. Delbrück escaped from it by switching to phototropism of the fungus *Phycomyces*, hoping again that simple experiments and clear thinking would lead him to a breakthrough. Twenty years of work, however, failed to bring the solution of this very difficult problem any nearer.

Fischer's biography reveals Delbrück as a German Romantic searching for the



Double Max — Max Perutz (right) with Delbrück at a birthday party for Linus Pauling.

Holy Grail, which for him was Bohr's "Elementary Fact of Life". To those like myself, who have tried to understand the workings of large biological molecules in terms of simple chemical laws, Bohr's and Delbrück's belief in some mystical principle looks like vitalism, but this book has led me to understand the motive behind Delbrück's proverbial and often misplaced scepticism of new work. For example, he objected to Beadle and Tatum's one-gene-one-enzyme hypothesis on the ground that it could not be falsified by experiment; he dismissed Lwoff's lysogeny of phage as a non-phenomenon; and he disbelieved Meselson and Stahl's demonstration of the semi-conservative replication of DNA. Fischer writes that Delbrück wanted to model himself on his two great teachers by combining Bohr's insights with Pauli's mordant criticism, or, as he put it, by becoming God and Mephisto all in one. But I have the impression that Delbrück really wanted to disbelieve any advance that removed the elusive "Elementary Fact" further from his grasp. □

Max Perutz is at the Medical Research Council's Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

Explanation for the amiable physicist

Horace Freeland Judson

The Problems of Biology. By John Maynard Smith. Oxford University Press: 1986. Pp. 134. Hbk £12.95, \$19.95; pbk £4.95, \$6.95.

WHAT are "the problems of biology"? The notion is ambiguous. We could mean the next unanswered but perhaps accessible large questions — in effect, the list we would offer of the most promising lines of research on which a graduate student in biology might found a prize career. Or we could mean the approaches and constraints that distinguish biology from other sciences — in other words, the considerations we would explain to an amiable theoretical physicist who asks (as some still do) why biology doesn't provide real answers, why the ratio of theory to data seems so inordinately low. The latter sense of the problems provoked Ernst Mayr to the declaration of biology's epistemological independence with which, four years ago, he opened *The Growth of Biological Thought* — three chapters that together ran to nearly twice the length of the entire book in which John Maynard Smith now marshals the problems of the former sort, the questions biologists are getting on with.

Yet the semantic fork has a handle. The fundamental problem of biology in either sense is the presence, in systems of immense complexity, of the element of the contingent — of the accidents of history. That is, in everything that is to be explained, whether we are considering, say, the nature of mutations in regulatory genes, or the behaviour of a given mammal, or the array and interrelationships of all species surviving at a given moment, we confront phenomena shaped by uncountably long sequences of past events, many of them random. Contingency characterizes life from its origin. Darwin put history into biology and we're still working out the consequences.

Contingency, for many biologists, is normally an aspect of the background; but for the biology Maynard Smith practises it's up front. He is, after all, a population geneticist (and the recent target of a *Festschrift*) who has written a long-running paperback on evolutionary theory and who has in this past decade been modelling theorems of sociobiology as simple games. He says he agreed to write this new book for two reasons. He was flattered by the opportunity to

● Also published in this series (OPUS) are *The Problems of Evolution* (by Mark Ridley) and *The Problems of Chemistry* (by W. Graham Richards).

measure himself against Bertrand Russell's *The Problems of Philosophy*. More important, he had been brought up on "the popular books of Eddington, Jeans, Einstein, Haldane and Wells" to believe that "the fundamental ideas of science can be explained to anyone willing to make the effort needed to understand them". *The Problems of Biology* has most of the virtues of that splendid tradition. It is sensible, colloquial, bracingly lucid, and stripped, sometimes daringly, to essentials. The comparison that sprang to my mind is Lawrence J. Henderson's *The Fitness of the Environment* — a classic (incidentally American) that pre-dates that (chiefly British) line Maynard Smith recalls. One rises from each of these lean volumes with a sense of the world not merely clarified but put in order; one leaves the table with an appetite.

The definition of life is not now a real problem, as it was even four decades ago when Erwin Schrödinger's *What Is Life?* was attracting physicists into biology; but Maynard Smith takes it as an obligatory starting point. He re-states the two characteristics that we have come to recognize as definitive. First, living organisms, always in populations, reproduce, and not simply by multiplication — as does a fire — but by means of a hereditary mechanism such that like begets almost-like and the population evolves individuals successively more adapted for survival. Secondly, an organism is "a complex structure which is maintained by the energy flowing through it". Maynard Smith calls these two aspects of what we want to understand about any living creature or process the "ultimate and the proximal causes", adopting terminology that Mayr, for one, used. Ultimate causes are evolutionary, proximal causes functional in the sense that they answer the question "how?" in terms of immediate process. Heredity, which in the narrow sense of genetics has become "perhaps the best understood and least problematic area of biology", Maynard Smith considers to be the necessary prelude to the cluster of ultimate problems: the evolutionary process, the origins of sexual recombination, the levels of life and the patterns of nature (the species and their interactions) that have resulted from evolution. He starts the proximal problems with the organism's methods of achieving stability and controlled responses at the biochemical level, and goes on to animal behaviour, neurobiology, and the development and differentiation of the organism. He gives his tenth and final chapter to the origin of life. The list is canonical, the particular questions within these categories incisively posed, the examples apt.

Yet a book both brief and universal also requires an organizing principle. The right one was to hand in the all-pervasive

contingency of biological phenomena. Maynard Smith fails to exploit it. The argument runs along the following lines. The unsolved proximal topics, to begin with, are intimately related. As he notes, two are chief among them: development and behaviour. At least since William Harvey and the realization that *omne vivum ex ovo*, development and differentiation has been the great intractable problem. We now see that it subsumes, for higher organisms, the problem of the controls on biochemical activity in the individual cell — for these, directly or at one remove, are special cases of gene expression. We see that the most extreme elaboration of development is presented by the differentiation of the immune system (a problem of the first rank, surely, but not addressed by Maynard Smith) and of the nervous system and brain. We now get the first glimpses of a relationship, at the level of receptors common to lymphocytes and nerve cells, between the immune and nervous systems, whose mutual interactions are giving rise to an unexpected new field, behavioural immunology. We are reminded that receptors on neurones are to be understood — Jean-Pierre Changeux has been pressing this point — as allosteric proteins, which, of course, were first recognized as control elements within the cell. The essence of allosteric proteins is that they permit networking of responses free

of chemical stringencies. All this suggests that the problem of development even implicates the problem of behaviour, so far as that can be approached from the bottom up.

Development and its subsidiary problems are as difficult as they are for at least two reasons: the individuals and systems in which we are most interested are complex beyond imagining, perhaps beyond describing, and the vast blueprint for what we encounter as this proximal complexity is itself a palimpsest, layer upon layer upon ancient layer of alterations each of which could only arise directly from what was available immediately before. In evolutionary biology the most contentious problems are the nature of the changes that have led to speciation and the rates at which evolution has taken place, along with the special case of the evolution of animal and human behaviour. The unifying relationship here returns to the control genes that determine development, in particular how these mutate, perhaps accounting for bursts of evolutionary change. To fathom the proximal we require the ultimate, and the other way around: biology is intrinsically, in part, *ad hoc* — as we were explaining to the amiable physicist. □

Horace Freeland Judson is Professor in and Director of the Program in Science and Writing, Johns Hopkins University, Baltimore, Maryland 21218, USA.

Taking the tablets

Michael Nelson

How to Live Longer and Feel Better. By Linus Pauling. W.H. Freeman:1986. Pp.322. Hbk \$15.95, £15.95; pbk \$7.95, £5.95.

AN ELDERLY relative from America visited me recently in London. At breakfast each morning she swallowed a vast array of vitamin tablets, one "for cholesterol", one "for colds", one "for skin" and so on. "Do you feel any better for taking them?" I asked. "No", she replied, "but I think I ought to and they don't seem to do me any harm".

In *How to Live Longer and Feel Better*, Linus Pauling uses equally unconvincing arguments to try and convince readers of the benefits of the megavitamin treatment of life. He begins by describing "The Regimen" for better health, outlining in declamatory fashion his reasons for consuming anywhere between 10 and 3,000 times the recommended intakes of most known vitamins, with emphasis on the particular value of vitamin C. He then discusses the roles of protein, fat, carbohydrates, energy, water (and, of course, vitamin C) in maintaining health, and

provides some sensible if unoriginal ideas about balanced eating and avoiding fad diets. His emphasis on the protective role of vitamins, however, is at odds with current epidemiological evidence regarding the effects on health of smoking, excess weight and lack of exercise. This first section also contains a number of irritating nutritional inaccuracies (for example that starch is found in all fruit, that vegetarian diets cannot provide the correct balance of essential amino acids, that seeds and nuts are low in vitamins).

Dr Pauling regards vitamins as food, and so makes the absurd suggestion that "It is wise to supplement the vitamin supplements with a good intake of fruits and vegetables" (p.31) — this is standing nutrition on its head. Also, I am instinctively wary of advice which tells me that "swallowing not more than a half-dozen tablets a day" isn't burdensome, and find misguided the remark that a regimen adding £7 to a weekly food bill is inexpensive.

The second section, "The New Nutrition", begins with promise. A history of the discovery of vitamins is followed by an interesting chapter on vitamins and evolution. Unhappily, this degenerates into an attempt to establish optimal levels of vitamin C intake for human beings based solely on comparative physiology. Anal-

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

"ACTUALLY I STARTED OUT IN QUANTUM MECHANICS, BUT SOMEWHERE ALONG THE WAY I TOOK A WRONG TURN."

gies with pigs, cats, monkeys and goats provide interesting but unconvincing evidence for the optimal human requirement, and direct evidence from experiments on healthy people is sorely lacking.

The third section on orthomolecular medicine is, frankly, tedious. There is very little here that has not already been said elsewhere. For illness after illness, from the common cold to AIDS, Dr Pauling presents repetitive and wholly one-sided arguments in favour of treatment with massive doses of vitamin C. Those few studies quoted which do not support his contentions are dismissed for having failed to provide sufficiently large and lengthy doses to be effective. This lack of balance lends an evangelical tone to his writing and casts justifiable doubt on his interpretation of others' findings. For example, Knox's data on deaths from heart disease (*Lancet*, **1**, 1465-1466; 1973) did not, as Pauling claims, show an association with vitamin C "greater than that for any other factor" (p.151). The correlation between regional mortality from heart disease and diet was greatest for calcium ($r=-0.67$), not vitamin C (-0.49), and in a multiple regression analysis, calcium, fat and vitamin D — not C — were the most significantly correlated variables.

This book, then, is for the lay audience which wants to be convinced that life will be happier and longer if Dr Pauling's prescription is followed. It is not for epidemiologists or nutritionists who are looking for reasoned arguments on the merits of vitamins in the prevention of disease. There is a comprehensive bibliography of interest to those who wish to look into the subject themselves and draw their own conclusions. □

Michael Nelson is a Lecturer in the Department of Food and Nutritional Science, Kings College London (KQC), Campden Hill Road, London W8 7AH, UK.



New Books

TEMA

Conference Proceedings of the International Symposia on "Trace Elements in Man and Animals" at Aberdeen from 2-6 July 1984.

Published March 1986, 950pp,
ISBN 0 85198 533 X
UK £43.00 Americas US\$81.70
Rest of World £47.30

Tylenchida, Parasites of Plants and Insects

by M. R. Siddiqi

Published 1986, 512pp, 150 illus.,
ISBN 0 85198 554 8

A basic text and guide to the identification of the largest and economically most important genera of plant-parasitic nematodes.

UK £45.00 Americas US\$85.50
Rest of World £49.50

Sugarcane Diseases

by A. Sivanesan and J. M. Waller

Published 1986, 88pp, 36 figs
ISBN 0 85189 564 5

An up-to-date account of sugarcane pathogens with synonymy listed and including taxonomic groupings and identification keys.

UK £12.50 Americas US\$24.00
Rest of World £13.75

International Savanna Symposium 1984

Edited by J. C. Tothill and J. J. Mott

Published 1985, 38pp,
ISBN 0 85198 535 1

Held in Australia in May 1984, attention is directed at the similarities and differences in the ecology, land use and management of the world's savannas.

UK £25.00 Americas US\$47.50
Rest of World £27.50

The Relascope Idea

Relative Measurements in Forestry

by Walter Bitterlich

Published 1984, 256pp, 118pp of illus.,
2 col. pl. ISBN 0 85198 539 4

From the theory of angle-count sampling and the development of relascope instruments to the Telerelaskop, and its acceptance as an accurate dendrometer.

UK £28.90 Americas US\$58.00
Rest of World £31.80

For orders, enquiries and further information on these or other CABI books contact:

C-A-B International, Central Sales,
Farnham House, Farnham Royal, Slough
SL2 3BN, UK
or

C-A-B International,
North American Office,
845 North Park Avenue, Tucson,
Arizona 85719, USA

Reader Service No.37

Developing a sense of the past

D.R. Newth

A History of Embryology. British Society for Developmental Biology, Symposium 8. Edited by T.J. Horder, J.A. Witkowski and C.C. Wylie. *Cambridge University Press*: 1986. Pp.477. £60, \$99.50.

WE MAY reasonably date the birth of modern embryology from the middle years of the last century, and name von Baer as the midwife. It was then, after all, that the words embryology and embryologist entered the language, and then that embryologists accepted a share in the eminently practicable programme of comparative anatomy. Both comparative anatomy and descriptive embryology were, of course, soon to be swept up in the Darwinian revolution, and many embryologists welcomed their new role as seekers for recapitulated phylogenies and were content to wait for the expected reward, an explanation of just how the events of ontogeny could be determined, in sequence, by information from numberless generations of adult ancestral forms. In his essay in this volume, Ridley charts the history of post-Darwinian embryology in Britain, and brings to life the personal and scientific factors shaping it between the time of Balfour and the 1930s.

Other embryologists preferred to tackle the nature and control of developmental processes more directly, with the methods of *Entwicklungsmechanik*. For them cell theory and genetics provided the real boundaries of the area in which theories of development could usefully operate. A hundred years after experimental embryology started is, perhaps, a suitable time to try to assess its achievements and failures.

Despite the title, this symposium volume does not constitute "a history of embryology", nor even its history over the past hundred years. Rather, in 15 papers linked or introduced by editorial comment, it provides something of the flavour of the period and critical accounts of the work and attitudes of some of its leading figures. Of the contributions that deal

mainly with individual scientists (Weismann, Boveri, Morgan, Harrison, Spemann and Waddington) the treatment of Hans Spemann by Horder and Weindling is the most ambitious. Spemann's scientific life is related both to the embryological and to the academic and social environment in which he worked. This gives us a satisfying complement to the impersonal history of the primary inducing factor written by Saxen and Toivonen and the relevant parts of the more personal account by Jean Brachet of the beginnings of biochemical embryology.

The fact that inductive and organizer phenomena were discovered and investigated by embryologists, and were clearly biologically important, gave attempts to explain them in cellular and biochemical terms a symbolic value for embryology. The disappointment when early hopes were not realized persists to this day and colours most comments on the present status of organizer work. The dispersal of the Spemann school by Nazi politics and by war was not the real trouble, it only exacerbated an existing crisis. One suspects that this crisis, as much as anything, prompted Medawar's reproach that "there has therefore been little sense of progression or timeliness about embryological research". Forty years hence, in retrospect, it may be possible to take a more cheerful view, but the central position once held by induction is unlikely to be restored.

To respond to an inductive stimulus a cell must both be competent and in the right place (that is, in contact with the inducer). But cell position, if known to the cell, could alone be a sufficient cue to appropriate developmental behaviour, as Driesch first proposed for highly regulative systems. The idea has since had a varied history, which is lucidly surveyed by Wolpert whose own formulation of the concept of positional information is free from the imprecisions of some of its predecessors and from the deficiencies of others.

Of the contributors dealing with the relations between embryology and genetics, two make it their sole theme. Allen's account of T.H. Morgan sees him as an important agent in the creation of a still-unhealed split between the two, but not as

The book

*We come from halls of Academe
And, on an inspiration,
Alone, or in a learned team,
Embark on publication.*

*We seek to teach, or disabuse,
Where controversy rages,
And press our honest facts, or views,
On academic pages.*

*Then, if our theories endure,
Maintained by erudition,
Our luck and virtue should assure
A third or fourth edition.*

*For ultimately all should know
The fruits of our endeavour.
Though scholars come and scholars go,
Our books go on for ever.*

Ralph A. Lewin

the maker of that split; indeed, given the nature of the material in each case, some divergence was inevitable. Morgan himself always conceded that a complete genetic theory had to come to terms with the problems of embryonic development even if it was not designed to solve them. Sander, in his contribution, takes the split as his starting point and explores the interface between genetics and insect development that has nevertheless arisen and proved very fruitful.

This volume shows that the legacy of Needham in Britain and of Oppenheimer in the United States is being built upon by a new generation, some working embryologists, some historians of science and some philosophers. They are finding plenty to interest them in twentieth-century biology where the sources are, naturally, far richer than for earlier times. Only

the very closeness of their material will embarrass them, and not only for the reason that Sander mentions: the risk of ruining friendships of long standing "when the deeds and thoughts of living persons are involved". All embryologists and many other experimental biologists will be glad of this book for itself, but also for the promise it implies of further historical work to come. I have only two complaints. Sooner or later, and better sooner than later, we shall have to convert lip service to unifying trends in biological science into action by taking plant material into consideration. But if this volume lacks plants it certainly does not lack misprints. □

D.R. Newth, Monevechadan, Lochgoilhead, Cairndow, Argyll PA24 8AN, UK, was formerly Regius Professor of Zoology at the University of Glasgow.

Terms of medicine

D.J. Weatherall

International Dictionary of Medicine and Biology. Editor-in-chief Sidney Landau. Wiley:1986. Three volumes, pp. 3,200. \$395, £285.

IN HIS excellent biography of Samuel Johnson, John Wain quotes a telling passage that sums up that remarkable man's emotions on completing the *Dictionary of the English Language*, published in 1755: "I have protracted my work till most of those whom I wish to please have sunk into the grave, and success and mis-carriage are empty sounds. I therefore dismiss it with frigid tranquillity, having little fear or hope from censure or from praise". Although less self-revealing than Johnson, Sidney Landau, in his editor's note which introduces the *International Dictionary of Medicine and Biology*, leaves the reader in little doubt that, despite the advantages of computerization and a large committee of co-editors, the lexicographer's lot has not changed much over the years.

The sheer statistics of this new endeavour are staggering. The *Dictionary* contains more than 151,000 terms and more than 159,000 definitions (I have taken the editor's word for the accuracy of these and the other figures). Its computerized database consists of about 27.5 million keystrokes or characters, corresponding to 22 million printed characters or about 30 books of a standard length. The editor-in-chief has been helped by 81 advisory editors, 10 co-advisory editors and 90 contributors who dealt with the definition of more than 70 subjects. About two-thirds of the advisory editors are American, the rest British. The total gestation period was 11 years.

The object was to cover the traditional medical sciences, and, in addition, those areas of the biological sciences which relate to the practice of medicine. The collection is not limited to current usage, though terms which were obsolete by the time of the First World War are excluded unless of particular historical interest. The *Dictionary* also includes the derivation and alternative usage of every term defined. Usage is American throughout; British spelling is occasionally cross-referenced (e.g. for *foetal* see *fetal*).

What is the reviewer to do, faced with over 3,000 pages of lists of medical terms and biological potpourri? Pope's instructions to the critic to "survey the Whole, nor seek slight faults to find" never seemed more apposite. And, viewed in this way, the whole is very impressive. The definitions are clear and unambiguous, the nomenclature consistent and the method of cross-referencing, an enormous problem in work of this size, very effective. The volumes are well produced and built to last, and the print is clear and attractive. The scope is enormous. Syndrome lovers will have a field day, with 43 pages containing nearly 800 entries among which are some genuine collectors' pieces such as the *whistling face—windmill vane hand syndrome*. There are even flashes of relief, not always intentional one suspects. After many exhausting pages which list every conceivable human muscle the next entry is *muscle-bound*, defined as *possessing an excessive physical musculature that is functionally suboptimal*. But despite this mass of traditional medical jargon, the work is remarkably up to date, containing as it does much of the vocabulary of molecular medicine.

It is only when the reader, forgetting Pope's advice, examines a particularly familiar topic in more detail that some of the problems of a compilation of this type become apparent. For example, under the

heading *Anemia* there are nearly 200 entries. Many are obsolete terms, and at least some of them cross-reference to even more obsolete ones which surely cannot be even of historical interest. And there are some odd entries. *Molecular anemia* for example. I doubt if this term is (or was) used. Here, it is defined as *anemia due to an abnormal hemoglobin molecule*. Presumably the confusion is with *molecular disease*, a term which was used by Pauling and Itano to describe sickle cell anaemia in their classical paper in *Science*. And I doubt if many physicians would have much cause to search for *febrile pleiochromic anemia* which is cross-referenced as an alternative to *thrombotic thrombocytopenic purpura*. In some cases it is made quite clear that terms are obsolete or rarely used; in others, such as this one, the reader has to guess. Also, presumably in an effort to be all-embracing, there are some rather gratuitous entries; even allowing for the current state of the health services on both sides of the Atlantic, it is doubtful if many will have occasion to look up *therapeutic pessimism*.

Was all the effort worth it? I think so. With the extraordinary breadth and diversity of the medical sciences there surely is an honoured place for a dictionary such as this, one of genuine scholarship which provides a comprehensive coverage of the language of medicine. Undoubtedly work should start on a second edition. But whether the publishers will be able to find editors to take on this onerous task is another matter. Sidney Landau's editorial note, though low key, lets slip a few of his problems. In discussing the question of obsolescence he points out that the rate of change and continuity of the past varies considerably between specialties; the kind of terms that an immunologist and an anatomist identify as obsolete are very different. Anyone who has sat on a curriculum committee in a medical school will feel for him.

Those who suspect that the medical profession has spent several thousand years developing a language with which to impress its patients and to confuse the world at large can now rest their case; the evidence is set out unambiguously in these massive volumes. But this grand obscurantism still has its advocates. Surely, they would argue, those unfortunate enough to have measles involving the lung might feel just that bit better, or at least impress their friends or, perish the thought, their bank managers, if they arrived home with the label Swyer-James-Macleod syndrome (or the Swyer-James syndrome; or the Macleod syndrome; or, best of all, the Swyer-James unilateral hyperlucent lung syndrome). □

D.J. Weatherall is Nuffield Professor of Clinical Medicine at the University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

View to prevention of cancer

John Cairns

Cancer: Risks and Prevention. Edited by M.P. Vessey and Muir Gray. *Oxford University Press: 1985. Pp.301. £20, \$27.95.*

Cancer: An Enigma in Biology and Society. By R. Nery. *Croom Helm, London/The Charles Press, PO Box 15715, Philadelphia, Pennsylvania 19103, USA:1986. Pp.436. £29.50, \$32.50.*

UNTIL the 1940s, cancer research was almost entirely in the hands of pathologists and surgeons. You went to the former for a precise diagnosis and a statement of your prospects, and to the latter for treatment. Although many scientists were struggling to understand carcinogenesis, for example by studying the structure of the different carcinogens and looking for peculiarities in the metabolism of cancer cells, textbooks on cancer remained essentially textbooks of pathology.

Recently all this has changed. The epidemiology of cancer has come under ever closer scrutiny, and as a result we now see that cancer is just another of the preventable diseases — perhaps just as preventable at the end of the twentieth century as were the infectious diseases at the end of the nineteenth century. The treatment of cancer, in particular the development of chemotherapy, has become a multi-million dollar industry inhabited by an army of industrial chemists and genetic engineers (making new products), clinicians (testing the products) and experts in public relations (extolling the virtues of it all). Finally, we are beginning to understand the underlying biology of cancer in terms of certain alterations in genes that affect the exact chemistry or level of expression or subsequent modulation of certain gene products.

This great diversification of cancer research has made it very difficult to write an up-to-date book to serve as a vade-mecum for someone entering the field. There is so much explaining to be done, so much background information to be incorporated which inevitably breaks up the sequence of material that should go into each chapter. How, for example, could one design a good account of the natural history of cervical cancer that would not be thrown completely out of balance by the long dissertation on DNA hybridization

that is required if the reader is to understand what we now know about the causation and diagnosis of the precancerous conditions of the cervix? To add to these difficulties, cancer research is not a real scientific discipline. Admittedly, it has become more and more entangled with semi-hard sciences such as biochemistry, molecular biology, developmental biology and even statistics, but unlike those subjects it still seems unrelievedly idiosyncratic because it does not build steadily upwards from a firm base. We can be sure that textbooks of biochemistry, written 100 years hence, will include descriptions of DNA structure and intermediary metabolism that are quite like the descriptions of today. But the future textbook on cancer will probably be largely unrecognizable, save for the sections dealing with the epidemiology and crude anatomical classification of the various cancers. Interestingly, these are two subjects that can still be described in relatively non-technical terms. For example, although epidemiologists employ complex computational methods and sophisticated forms of analysis, the end-results of their efforts can be described using words that could surely be understood by almost anyone.

Partly for these reasons, *Cancer: Risks and Prevention* is a joy to read. It is a set of essays, assembled as a *Festschrift* for Richard Doll on the occasion of his retirement from Green College to a career in "full-time cancer research" (these are the words of the editors). I do not know of any really satisfactory textbook on cancer epidemiology. This is no exception, mainly because it does not include a dissertation on the various techniques used in the practice of epidemiology. The editors point out in the preface that what they have assembled is not meant to be a comprehensive text. Nevertheless, until something more complete comes along, it and the monumental article by Doll and Peto on the causes of cancer, published in the *Journal of the National Cancer Institute* (66, 1191–1308; 1981), are the works I shall recommend as essential reading for everyone entering cancer research. The chapters cover all the main topics — diet, smoking, radiation, occupation, infective agents, hormones, drugs and even the use of legislation in the quest for prevention — and they have a certain homogeneity of excellence that is the sign of skilful editing. Taken as a whole, the book gives a wonderfully clear and exciting picture of the field. After reading it I feel more than ever that we will eventually learn how to prevent most forms of cancer to the extent we now prevent the major infectious diseases. This, much more than the breathless accounts of agents such as interleukin-2 which are currently making the headlines, is surely the real message of hope.

The other book, *Cancer: An Enigma in Biology and Society*, is a much less felici-

tous work. Somehow it seems almost perversely old-fashioned. The highlights of contemporary cancer research are the epidemiology, the molecular biology and the increasing use of carefully designed trials to study treatments and possible means of prevention. None of these subjects is dealt with satisfactorily by Dr Nery. Instead, he tends to dive into the distant past at every opportunity. Tracing



Eye for epidemiology — Richard Doll in 1969, the year of his appointment as Regius Professor of Medicine at Oxford University.

notions about the nature of cancer back to the "spagyric" theory of the sixteenth century may be a mark of erudition but, these days, such learning has a tragico-comical air to it. No amount of erudition can make up for the absence of an up-to-date description of the relationship of oncogenes to chromosomal rearrangements. One of the most exciting developments in the past five years has been the identification of the precise genes whose functions or level of expression is altered by the various chromosomal alterations seen regularly in certain cancers; yet all this gets hardly a mention, and once Dr Nery has given the pathologist's view of the common and less-common chromosomal changes seen in cancer cells, he quickly retreats to atrabilism in ancient Greece and Islam. Plainly, he is uncomfortable with the developments of the past ten to twenty years.

I do not want to seem too unevenhanded in my review of these two books. The totality of cancer research is not an easy subject for one person to deal with, and it does not fall naturally into a single book. One day, it will. But, for the time being, the best reading is bound to come from more limited overviews, especially accounts of the epidemiology and the molecular biology. □

John Cairns is in the Department of Cancer Biology, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA.

• A new volume in the series *Readings from Scientific American* is *Cancer Biology*, with introductions by Errol C. Friedberg. The book includes reprints of thirteen articles by prominent research workers (among them John Cairns) which cover "much of the essential core" of the subject. Publisher is W.H. Freeman, price is pbk \$12.95, £12.95.

Killing creatures with kindness

Alan Holland

The Case for Animal Experimentation: An Evolutionary and Ethical Perspective.

By Michael Allen Fox. *University of California Press*: 1986. Pp. 262. \$18.95, £15.95.

THERE can be few people who have not benefited in one way or another from the results of experiments on animals. For this reason it seems unjust that researchers alone should bear the brunt of the hostility directed at such experiments. By the same token, however, if they are to retain public support for their endeavours, scientists have to accept the legitimacy of public scrutiny and ensuing legislative restrictions.

Those whose work involves animal experiments will be grateful to Michael Fox for his timely and valuable contribution to a debate which runs the risk of degenerating into a clash between vested interests and pressure groups. His aim is a "rational compromise" between the scientific community and animal protectionists, and his chosen brief is the defence of animal experimentation in biomedical and behavioural research. Although this restriction of ground, taking in about 40 per cent of the whole, is no doubt reasonable, it does little to allay a general sense that the cause of experimentation is fragmenting under pressure. Lines are being drawn between the interests of industry and those of academia: research whose results are available for the benefit of all is more easily justified than research whose results, for political or economic reasons, are not open to scrutiny. Moreover, biomedical research, which is largely carried out on anaesthetized animals, is more easily defensible than behavioural research which mainly involves alert and conscious animals.

Two aspects of Fox's position illustrate nicely the obstacles in the way of rational compromise. One concerns basic research. Many researchers seem to accept that their work must ultimately be justified in terms of likely medical benefit. Fox demurs, taking the prospect of a significant advance in knowledge as justification enough, provided only that it "may conceivably" contribute to the alleviation of

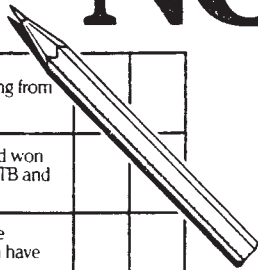
suffering (p. 197). He is right to press the point. If the deliberate imposing of suffering is ever justified, it is far from self-evident that the alleviation of suffering is the only value that can do so. The other obstacle concerns the general agreement that it is desirable to avoid unnecessary suffering; Fox spells out with admirable rigour what this really means (pp. 165–168). But he proceeds to subsume under this head a caveat against causing *excessive* suffering, thereby construing "excessive" not in relation to the sufferer but to the likely benefits. On his account, therefore, there is no automatic cut-off point in the imposition of pain, if the prize is high enough. This conflicts with the recommendations of the Canadian Council on Animal Care, which are helpfully listed in an appendix.

Fox derives his guiding principles from an examination of the ethical and evolutionary relationships between man and animal. He holds, as did Kant (who is clearly his inspiration here), that there can be no moral objection as such to humans treating animals as means to their own ends. He bases this view, first, on a relational and anthropocentric theory of value, according to which, in the Universe as we

know it, human beings are the source and reference point of all value; second, on a denial that animals have moral rights, on the grounds that such rights can only develop in the context of a community of autonomous beings. On the other hand he holds that an animal's capacity for suffering, along with certain other features, makes it an appropriate object of moral concern and precludes cruel or inhumane treatment. Fox's reading of the evolutionary relationship between man and animal vindicates his ethical stance, but only thanks to his somewhat disconcerting rejection of Darwin's view that the difference is "one of degree and not of kind". Attention to the work of the "evolutionary epistemologists" (K. Lorenz, R. Riedl, D. Campbell) would, I surmise, yield an evolutionary perspective less congenial to his standpoint.

On the basis of these principles, Fox proceeds to show that most behavioural and biomedical experiments are morally acceptable. Through a series of illuminating case descriptions he illustrates also the important difference between retrospective and prospective assessment of the justification for an experiment. In describing the Harlows' experiments with rhesus monkeys which involved severe maternal and sibling deprivation, for example, he is prompted to doubt whether "these nightmarish and regrettable experiments" would "pass muster before today's ethics review committees" (p. 103). But he regards the results obtained as a redeeming factor. Retrospectively the experiments lie on the borderline of acceptability, but they might not have been licensed in advance by his criteria because of uncertainty over the outcome. By contrast, Fox has little difficulty in justifying all of the experiments at Toronto's Hospital for Sick Children that he describes, because all promise at least moderate benefits and none involves more than moderate suffering. In these accounts, Fox rightly stresses the importance of spelling out the context of the experiments in order to arrive at a fair adjudication of their worth. More important still he goes on to show how it is only by understanding the demands of scientific inquiry in general — the need for controlled experiments, the requirement of repeatability, the need for statistically significant results — that we can avoid misconstruing a great deal of experimentation as gratuitous. His extensive review of the alternatives to the use of animals in research

ANIMAL EXPERIMENTS IN MEDICAL RESEARCH YES OR NO



Would you treat a child suffering from leukaemia?		
Would you retain Society's hard won control over polio, diphtheria, TB and smallpox?		
Would you agree we must have medicines and vaccines which have been tested for safety?		
Would you agree that we need to alleviate and control, for example, cancer, arthritis, multiple sclerosis and heart disease?		
Would you like to see a cure for AIDS and Legionnaire's disease?		

Animal experimentation has made an essential contribution to the control and eradication of serious diseases. Much more requires to be done - this work must continue.

THE RESEARCH DEFENCE SOCIETY

Safeguard your future

RESEARCH DEFENCE SOCIETY, GROSVENOR GARDENS HOUSE, GROSVENOR GARDENS, LONDON SW1W 0BS
Researchers' response: a recent advertisement in a British newspaper.

leads him to consider that the aim of abolition, even in the long term, is unrealistic. Indeed, by his principles, there need be no moral objection at all to experiments on anaesthetized laboratory-bred animals continuing indefinitely.

Should we accept his principles? There is no doubt they will be hotly contested. His suggestion that man is something more than an animal is distinctly odd and even demonstrably misguided in so far as it rests on the allegedly paradoxical idea that man is both "part of" nature and yet stands "outside" it (p. 19). For there is surely no paradox in supposing that man belongs wholly within nature while at the same time having the capacity to reflect on his position. Again, Fox's crucial claim that animals lack intrinsic value seems to rest on the unsupported proposition that only beings capable of *assigning* value can *have* intrinsic value. What his theory of value does support is that animals only have intrinsic value if humans assign it to them. The most he can then argue is that they cannot be assigned an intrinsic value equal to that of humans.

Although marred occasionally by unworthy and on the whole irrelevant attacks on animal welfarists for being misanthropic, sentimental or irrational, a sincerity of tone runs through this book, nowhere more so than in the brave but unsuccessful attempt to reconcile the practice of experimentation with the claim to be humane; unsuccessful, because being humane implies being kind, and the deliberate infliction of avoidable suffering on an animal is not compatible with being kind to it. We must simply face the fact that the attempt to alleviate human suffering may involve the inhumane treatment of animals. Nor can the point be turned by the oft-repeated but hollow claim that animals stand to benefit as much as humans do. For the life of me I cannot see why an animal should be supposed to be more excited about being sacrificed to alleviate the suffering of another animal than it is about being sacrificed to alleviate the suffering of a human being. □

Alan Holland is a Lecturer in the Department of Philosophy, University of Lancaster, Lancaster LA1 4YT, UK.

ENGLISH & AMERICAN ACADEMIC BOOKS BY POST

to all countries

S. BENARD

74 Rochester Road,
Aylesford, Kent ME20 7BJ,
England

Living alternatives

James E. Lovelock

Origins of Life. By Freeman Dyson. Cambridge University Press: 1986. Pp. 81. £7.50, \$7.95.

Origins: A Skeptic's Guide to the Creation of Life on Earth. By Robert Shapiro. Summit, New York: 1986. Pp. 332. \$17.95.

EVERY so often a great physicist is perturbed by a curiosity about the meaning of life and is drawn in from that remote, personal Oort cloud where cosmology and subatomic particles can be considered in splendid detachment. One such was Erwin Schrödinger, whose small volume *What is Life?*, published by Cambridge University Press in 1946, still circulates and enlightens; it was, I would say, the book that most influenced my own thinking about science. Freeman Dyson's *Origins of Life* is modelled on Schrödinger's book, and like it is beautifully and economically composed.

Robert Shapiro has written about an entirely different search for the origins of life. Shapiro is a chemist intimately familiar with the experiments and theories concerned, and he writes like someone who lived when chemistry had soul, when chemicals could be handled personally and when it was permissible to wonder about their strange odours and crystalline forms. His book, he says, is intended to be an explanation "of what science does and does not understand about how life first began".

Schrödinger is best remembered for having sown the seeds of molecular biology. The great success of that undertaking has tended to make us forget that it was only part of his message, for Schrödinger also drew our attention to the fact that life processes are a legal avoidance of the second law. He asked the right question, "What is life?", rather than the equivocal one, "What are its origins?". Ancient happenings like the origin of life can never be known in detail by ordinary physical means. A message travelling through space and time attenuates until it is indecipherable. There is evidence about the state of the Earth when life began, but only because the persistence of life itself provides a limited but noise-free channel of communication back to our original ancestor. The geological record of the rocks of those times is almost absent.

Like a message passed by word of mouth, the chain of life back to the remote past carries precise but inaccurate information. But it is the only way we have to conjure up what the origin might have been. Shapiro describes the alternative guesses through the eyes of an enthusiastic explorer, with a friendly sceptic as a companion. It is so well done that one feels

like a privileged visitor watching experiments in progress on the laboratory bench. Dyson goes further and provides a model of a pre-nucleic-acid chemical evolutionary system. He calls it his "toy" model, but it is full of insight as good and simple models should be.

The efforts to discover the origins of life may be as unrewarding as those to discover our own personal forebears. They resemble another recent scientific quest, one touched on by Shapiro, the search for life on Mars. Such expeditions start out looking for something imagined but not defined. We know there is no life on Mars; indeed we knew it long before the Viking spacecraft landed there. The value of such journeys comes indirectly. The discovery that Mars is not alive is the neutral background against which the Earth was more clearly seen to be a living planet. Similarly, the voyage into the past to the origins of life may never reach its goal, but the preparation for the expedition powerfully concentrates the mind on the nature and properties of life.

The two books are a constant reminder of how complicated life is, and that it has no definition. The successes of molecular biology are so beguiling that we forget the organism and its physiology. Schrödinger's disciples, who founded the church of molecular biology, have turned his wisdom to the dogma that life is self-replicating and corrects its errors by natural selection. There is much more to life than this naive truth, just as there is more to the Universe than atoms alone — grandmothers live and enjoy the shade of Lombardy poplar trees not knowing that they and the trees are deemed by this dogma to be dead. Looking in from the outside, Dyson sees homeostasis to be as important a characteristic of life as replication. I would go further and suggest that, as well as matter, life has bulk properties and needs to be considered on a planetary scale. It is so tightly coupled to its environment that the boundaries of life extend far beyond the organisms they enclose. A tree or a coral reef are both more than 90 per cent dead matter yet they are alive.

These two books are a continuation of the quest started by Darwin and Schrödinger. Implicitly, both are more about the meaning of life than about its origins and both deserve to be widely read. They are personal yet scholarly accounts of a difficult subject, and are wonderfully open in the way they illustrate how science is really done. They indicate an alternative view to the profitable but nonetheless narrow orthodoxy of molecular biology. Not least they are the best possible answer to the mischievous myths of Creationism, even though that was not their first intention. □

James E. Lovelock, Coombe Mill, St Giles on the Heath, Launceston, Cornwall PL15 9RY, is an independent scientist.

Reflections from an interior world

Stephen Jay Gould

Memoir of a Thinking Radish: An Autobiography. By Peter Medawar.
Oxford University Press: 1986. Pp. 209. £12.50, \$17.95.

WITH the honesty and irreverence that have long been his trademarks, Peter Medawar begins his autobiography with three quotations, one his own, explaining why the lives of scientists make such dreadful reading. Choosing his own version from *Pluto's Republic*:

The lives of scientists, considered as Lives, almost always make dull reading... It could hardly be otherwise. Academics can only seldom lead lives that are spacious or exciting in a worldly sense... Their work is in no way made deeper or more cogent by privation, distress, or worldly buffetings... Academics lie outside the devastation area of the literary convention according to which the lives of artists and men of letters are intrinsically interesting, a source of cultural insight in themselves. If a scientist were to cut his ear off, no one would take it as evidence of a heightened sensibility.

Medawar's brave attempt in the light of his own strictures does, as he would readily admit, encounter these intrinsic problems of the genre. His life, as he describes it, has the continual interior excitement that is God's greatest gift to productive intellectuals, but not the exterior oomph that provides the stuff of good cinema. Moreover, where he has learned or experienced (as we all have) private items of greater import or taller tales, he has the unfailing decency to maintain both the trust and privacy of these knowledges. The central problem of autobiography is not confined to scientists: no person of integrity can possibly write with full disclosure, particularly for what we know of others (no matter how open we choose to be about ourselves).

Thus, one might say, at a superficial reading, that Medawar is hoist by his own petard. (I confess that I haven't the slightest idea what a petard is, but the phrase has such a nice ring. Update: I just looked it up — I always supposed it was a kind of tie; but it's an instrument of war, an explosive battering ram.) So hoist he is, but for one transcendent matter: Peter Medawar is so fine a writer (the world's best among scientists), so full of wisdom and insight, that he has been able to compose the very best example of what may be, by his own admission, a deeply flawed genre.

His *Memoir of a Thinking Radish* proceeds conventionally. (His odd, quintessentially Medawarian title, is a deliciously oxymoronic cut-and-paste job of Falstaff's and Pascal's epitomes for the human condition.) We learn of his misery tempered with good humour as a lonely boy at snobby public school (prep for American readers), where well-bred boys did not take

kindly to one "of Arabic extraction" (Medawar's father was Lebanese, his mother English; he quotes a friend of similar "extraction" who said: "I can't bear that expression, it makes you sound like some kind of gum".) In this section on his early life, Medawar first gives vent to a central theme of this book (so identified in his introduction): that Britain's greatness has been sapped not by loss of empire or



Three generations of Oxford teachers — members of the Theoretical Biology Club at Magdalen College in 1946. (From left to right) Francis Huxley, J.H. Woodger, Hans Motz, Karl Popper, J.Z. Young, Peter Medawar and Avron Mitchison.

other worldly things, but by the internal intolerance and petty narrow-mindedness of its ruling élites, a condition that he calls "snobismus", or simply "the English disease".

We then follow Medawar's career from Oxford, where he showed brilliance and general commitment to matters intellectual, but only found his specific path within science by probing, and trial and error; to years of fruitful research and stewardship of several academic and medical institutes (I had known of Medawar's seminal contributions to biological thinking, but not of his skills at, indeed strong fondness for, tasks of administration that I would find onerous or soul-destroying, but that he managed to invest with creativity); to his current life, after a serious stroke made bench-top laboratory work impossible, as a writer and director of research (and, in my view, as the biological guru of our age).

I particularly appreciated his candid descriptions and thoughtful generalizations about his primary work in immunology, especially his discovery of acquired immunological tolerance (the practical basis for techniques of grafting and transplant that have saved and extended so many human lives), and for which he and Macfarlane Burnet shared the Nobel Prize for Medicine and Physiology in 1960. Medawar emphasizes that the importance of these studies lay not so much in the specifics (for much of the work has been super-

seded both conceptually and practically), but in breaking an ideological lock that had subtly led scientists to a belief that such things could not be done. Many predecessors had not tackled the subject of organ and tissue transplantation because a prevailing view of genetic inexorability (stronger in more class-conscious Britain than elsewhere) had implied that a person's intrinsic construction could not be so compromised by accepting foreign bits into a working machinery. Medawar writes: "Some people had maintained that this was in principle impossible, since substances that provoke the normal rejection reaction are part of the genetic make-up, something that could no more be changed than one's blood group". And he con-

cludes, "the ultimate importance of the discovery of tolerance turned out to be not practical, but moral. It put new heart into the many biologists and surgeons who were working to make it possible to graft, for example, kidneys from one person to another".

Medawar has managed to transcend the limitations of autobiography by lacing his ordinary and largely anecdotal account with three saving graces.

First, he has provided practical examples from his career to illustrate the important themes that he has emphasized in his famous essays (see *Pluto's Republic*, *The Uniqueness of the Individual*, *Advice to a Young Scientist*). Medawar has shown, in these essays, how the conventional style of scientific writing misconstrues, even falsifies, the actual doing of science — a vital point for public communication since it is the procedures and methodology of science, much more than its content, that non-scientists so desperately misunderstand. The epitomized logic of inductivist accounts — from introduction, to materials and methods, results and conclusions — omits the basic human dimensions of hypothesis, confusion, error and collegiality. The false starts are in the wastebasket, not the *Science Citation Index*. By focusing his text on the frustrations, the errors and the bullheaded approaches, until kicked in the pants by data or good advice from colleagues with other perspectives, Medawar has illustrated his favourite

From *Memoir of a Thinking Radish*

themes by honestly discussing his own work.

Second, no one can match Medawar for verbal wit and dexterity; the *bon mot* on every second page compensates for any uninspired conventionality elsewhere. Where else can we learn so well about the dangers of Bible-reading for young children, the insipidity of American football or problems of defecation in a public school that instilled fortitude (and probably thought it inhibited homosexuality) by refusing to install doors in the lavatories: "This circumstance so offended the



From Memoir of a Thinking Radish

"Entire visual field is agreeably occupied" — Peter Medawar with his wife Jean in 1980.

Poet Laureate-to-be, John Betjeman, a Marlborough boy, that in his poem *Summoned by Bells* he implies that he had no bowel movement for three years".

Third, we find behind the witticisms, so well (and movingly) illustrated because implied rather than announced, that most rare and precious trait of courage, expressed, as an intellectual must, in word as well as deed. For more than a decade, Peter Medawar has worked, as productively as ever in his life, if not more so, under conditions of physical disadvantage (a major stroke and several aftermaths) that would have led most people to despair and resignation. Consider only his first words to his wife Jean, upon seeing her face after waking up from an operation (and fearing that a right-sided cerebral haemorrhage had destroyed vision in the left half of each eye): "My first words were 'Entire visual field is agreeably occupied.' I thought this remark apposite and well turned and it repudiated the case that my mind had deteriorated beyond hope of recovery".

The scientific autobiography is an irretrievably flawed literary genre, but when practised by the premier, the nonpareil, the numero uno, the top banana of the profession, it can be pretty darned good. □

Stephen Jay Gould is a Professor in the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

Resonance through the editions

John Maddox

A New Science of Life: The Hypothesis of Formative Causation (A New Edition).

By Rupert Sheldrake. *Anthony Blond, London: 1985. Pp. 278. Pbk £8.95.*

THOSE who review books are often faced with the task of dealing with books without merit; wise review editors advise that the appropriate treatment for them is no treatment at all, for even a scornful review of a second-rate work may seem like an advertisement for it. But what should be done about books that are worse than that, those that are perverse? Such documents are often designed to captivate opinion, and there is a case for denouncing them. But one should tread carefully. That, at least, is this journal's experience with Dr Rupert Sheldrake's book *A New Science of Life*, first published in 1981 and discussed in a leading article on 24 September of that year (*Nature* 293, 245; 1981) under the title "A book for burning?". What has happened now is that Sheldrake's book has been reissued in what is called a new edition, which differs most noticeably from the first in its bulky appendix filled with self-congratulatory references to the "controversy" following its first appearance.

Sheldrake, it may be recalled, is a biochemist turned metaphysicist who is the originator of the "hypothesis of formative causation", the assertion that the shapes of things, animate or inanimate, are determined by what are called "morphogenetic fields", by a process called "morphogenetic resonance". What seems to puzzle Sheldrake is that a journal such as this should be convinced that his book is a mischievous aberration. Here is an explanation, a routine argument so valuable in, for example, the return to their authors of manuscripts showing that special relativity is a pack of lies, or that quantum mechanics is mistaken, that it might be set up on a word processor for the more efficient conduct of editorial business.

The fact that there are unsolved problems within the framework of an existing theory does not of itself imply that the theory must be thrown away, or replaced by another; unsolved problems are the essence of science, the means by which theories are refined. Three centuries of physical science attest to that. In biology, the conventional theoretical framework is much younger; many would say that even Darwin's evolution became usable only in the 1920s, while molecular biology is just over 30 years old. Yet the Sheldrakes wring their hands about the puzzle that the difference between the genomes of differ-

ent species of *Drosophila* may be greater than that between the genome of human beings and the chimpanzee, shake their heads in disbelief at the prospect that the process of differentiation will ever be understood, assert (accurately, no doubt, but pointlessly) that there is no prospect of a "mechanistic" explanation of parapsychology — and plump instead for a theory which can account for everything, but at the cost of explaining nothing. So long as these morphogenetic fields have not been measured or otherwise described, their all-pervasive utility in the communication of form is merely a way of making accidents so generally reproducible that they become the rule.

This is elementary textbook stuff. Sheldrake's thesis as it stands is merely a simulacrum of a theory in the sense of an explanation, but for which Sheldrake claims support from Thomas Kuhn's innocent observation (in his valuable book *On Scientific Revolutions*) that people tend to resist the overthrowing of old "paradigms". Sheldrake takes this to be a licence for discarding any paradigm that comes to hand. His advocacy of his thesis is mischievous because it is overtly designed to give aid to that great company of people who think that the paranormal is the normal, that parapsychology is a more urgent problem for orthodox science than, say, that of differentiation in biology, and that Jung's collective unconscious is a fact of life.

Readers of this new edition will nevertheless observe that the past four years have not been conspicuously successful for formative causation. People have offered prizes for tests of the hypothesis, but the experiments seem to have been confined, for the time being, to the activities of television programmes. (A West German channel is said to be about to follow two British channels with a test of whether some people's knowledge of the hidden image in a puzzle picture will make it easier for others to tell the answer.) Lamely, the long appendix ends with the promise that "... other experiments ... are in progress in different parts of the world; and some of their results will probably be published within the next year or two".

The moral, for book reviewers, should by now be plain. Not all bad books should be censured, only those whose influence is so compelling, and so pernicious, that they must be resisted. Sheldrake's, which from the outset contained the ingredients of its own implausibility, should have been ignored. As things have turned out, the spurious controversy stirred in the appendix to this new edition gives the issue an air of durability it would otherwise have lacked. Presumably the appendix to the next edition will reprint this article, and so on.... □

John Maddox is Editor of *Nature*.

Bewitchment of the biologist

Philip Kitcher

Unfinished Synthesis: Biological Hierarchies and Modern Evolutionary Thought. By Niles Eldredge. Oxford University Press: 1986. Pp.237. £22.50, \$24.95

TOWARDS the end of his life, George Berkeley, famous Idealist and Bishop of Cloyne, wrote a discourse on the virtues of tar-water, in which claims about its curative power were mingled with lofty speculation on the main threads of Western metaphysics. Niles Eldredge's book recommends a cure for the ills of evolutionary theory, and, like Berkeley, Eldredge has a fascination with metaphysical abstractions. In both cases, the advertisements for the product seem somewhat overblown.

Eldredge's title introduces his principal theme and the central characters in his story. The modern synthesis is incomplete — in need of extension, he insists, not of replacement (pp.4 and 118). To complete it we must recognize the hierarchical structure of life, indeed we must see that organic nature is hierarchically structured in two different ways. Part of this necessary reconceptualization will involve changing the ontology of evolutionary theory to accommodate biological entities that have been ignored in the work of Dobzhansky, Mayr, Simpson and (especially) most of their followers (p.7). In this way, we shall be in a position to address some of the "additional topics" that the synthesis let slip from its field of vision — "such as the main events in the history of life" (p.206).

From the start we are plunged into the kinds of metaphysical thickets that give logical positivists nightmares. Eldredge argues, again and again, that the main deficiency of the modern synthesis is not to be exposed through the assembly of striking explanations of particular phenomena, familiar from the macroevolutionary studies of Stanley, Williamson and, perhaps most notably, Gould and Eldredge himself. Instead, the decisive argument against the synthesis lies in exhibiting its depleted ontology, specifically its failure to appreciate the "reality" of supraorganismic entities or, at best, its treatment of those entities as "classes" rather than as "individuals".

What does this mean? In my judgement, Eldredge has formulated his ideas and arguments using one of the least-promising suggestions in recent philosophy of science. Those who hail the thesis that species are individuals as a doctrine of revolutionary importance typically have

their attention fixed on a number of characteristics of species taxa: species are spatiotemporally bounded, they have "beginnings, histories, and terminations" (pp.144 and 158); species are held together by cohesion among their "parts", the "glue" in this case being the reproductive relations among constituent organisms — no sex, no species (pp.144 and 200). There is another suggestion, which I believe to be more significant, that Eldredge sometimes alludes to in passing discussions of the "individuality" or "reality" of an entity: genuine individuals "play an active role in the evolutionary process" (p.142).

If these are the important aspects of the individuality of species, then we should ask how appreciating them alerts us to the incompleteness of the synthesis. Were Dobzhansky, Mayr and Simpson logically compelled to deny that species have beginnings (to take the example where Eldredge has the strongest *prima facie* claim for incompatibility)? I think not. The criterion of reproductive isolation furnishes an objective principle for dividing an unbroken lineage — even if we are unable to apply it in particular cases — and the initial branching that gives rise to the lineage serves as a natural origin from which species taxa are objectively defined. Thus, even those who believe in gradual anagenesis do not have to deny that species have beginnings. (Histories, terminations and the reproductive nexus are easy.)

The real issue behind Eldredge's thesis of an unfinished synthesis is not a question of ontology at all. The dispute is about causation. Indeed, this is readily visible in Mayr's capsule presentation, which Eldredge quotes early in his book (p.5), and alludes to repeatedly thereafter: a main tenet of the synthesis is that macroevolutionary phenomena can be causally explained in terms of the known microevolutionary forces (primarily selection) acting on the known microevolutionary entities. *One* way to oppose this doctrine would be to argue that there are "higher-level individuals" that "play an active role in the evolutionary process" and to elaborate that point in terms of the importance of species selection. The debate would turn on both conceptual and empirical issues, but, thanks to Sober's analysis of causal concepts in evolutionary theory, and the studies of palaeontologists who have drawn inspiration from Eldredge, Gould and Stanley, we have some clear view of how it might be pursued. Instead, in *Unfinished Synthesis* Eldredge deploys (murky) philosophical arguments to turn his back on "multi-level selection" (pp.132, 137, 159) and hitches his wagon to a metaphysical cloud.

The patient reader who is prepared to penetrate the second-hand philosophy will discover insights of an imaginative

biologist. Eldredge starts with a discussion of four major works that inaugurated the synthesis, the first two editions of Dobzhansky's *Genetics and the Origin of Species*, Mayr's *Systematics and the Origin of Species* and Simpson's *Tempo and Mode in Evolution*. Although the stress on ontology sometimes distorts things, as in the attribution to Mayr of a schizophrenic view of species (pp.59 and 89), Eldredge provides a useful review of the explanatory projects undertaken and the explanations actually offered by the "founding fathers".

Similarly, in subsequent chapters, there are interesting *aperçus* of the later development of the synthesis and of the discoveries that are supposed to show the need for reform. Eldredge's (low) estimate of the importance of the debates about units of selection, about neutralism, even about punctuated equilibrium, is coloured by his disastrous fascination with ontology, but his comments on these debates are sometimes illuminating.

The main, original arguments in the book appear in the last two chapters. Eldredge uses his claims about "individuality" to suggest that there are two hierarchies of nested "individuals". The genealogical hierarchy comprises genes, chromosomes, organisms, demes, species and monophyletic taxa; the ecological hierarchy consists of molecules, cells, organisms, populations, communities and regional biotas. To understand evolution, we need to recognize the processes that link the levels in these hierarchies as well as the cross-hierarchy connections. Eldredge considers two extreme possibilities (pp.187–188), but his project is not that of formulating some definite version of evolutionary theory but of assembling the possible structures.

Why should anyone accept Eldredge's vision of evolutionary theory? *Unfinished Synthesis* provides two sorts of reason. The general argument from ontology looms large throughout, but I doubt that many scientists will find it compelling. Nor should they, for, despite the frequency with which Eldredge claims that certain theses are "obvious" or "forced on us" (see pp.89–91 and 104 for two among many examples), logically valid arguments cannot be generated by such simple means. Here the muddled metaphysics takes its toll, leading Eldredge to mistake a psychological conjunction for a logical tie.

The reason that is likely to appeal more to the empirically minded is to be found in the last 18 pages of the book. The new hierarchical view is used to account for two "specific evolutionary patterns": the prevalence of sexual reproduction and the progressive modification of surface "ornaments" in a trilobite lineage. In brief, Eldredge's answer to the first question is that species are "a simple and

necessary consequence of sexual reproduction" and that species are "difficult to dislodge" (p.200). His response to the second is that phacopine trends should be seen "in ecological context", with a prominent role given to "relatively major cross-genealogical, ecological" events that intensify extinctions (pp.211–212). Both explanations are provocative and interesting, but there seems no particular reason to attribute either to the power of hierarchical thinking or even to suppose that they are best articulated in Eldredge's preferred idioms.

For the issues are indeed questions about causation. What has sustained sexual reproduction against the apparently contrary force of individual selection (or genic selection)? What has caused the pattern we perceive in the part of the trilobite fossil record that Eldredge examines? The suggestions we are offered in *Unfinished Synthesis* seem not so much to be news — witness the kinship with Stanley's explanation of the persistence of sex and with the claims of Raup, Sepkoski, Gould and others for the importance of disruptions of ecosystems in the history of life; rather, they appear as forced translations of recent ideas, translations that impede the further development of intriguing, but incomplete, explanations. What causal forces does the synthesis (or contemporary orthodoxy) actually acknowledge? Do new proposals from macroevolutionary theorists compel us to introduce new forces? Questions such as these are at the heart of the current controversies about the need for a new evolutionary theory. *Unfinished Synthesis* does us the service of assembling some of the points around which discussion must proceed, but it does not analyse the issues with the precision, clarity and delicacy that they deserve.

There is undoubtedly grandeur in Eldredge's view of life, and the sense of grandeur has inspired some of his colleagues to hail it as a new departure in evolutionary theory. Others are likely to find the discussion grandiose and to miss the biological intelligence behind the metaphysical infatuation. In the 14 years since he and Gould offered novel ways of thinking about evolution, Eldredge seems to have mislocated what was most important in his contribution. Readers of this book may hope that he will return to the fertile ground that he helped first to explore. *Unfinished Synthesis* will aid the development of evolutionary theory if it encourages scholars to examine more closely the questions about causality that it assembles. □

Philip Kitcher is Director of the Minnesota Center for the Philosophy of Science, University of Minnesota, 355 Ford Hall, 224 Church Street SE, Minneapolis, Minnesota 55455, USA. His most recent book is Vaulting Ambition: Sociobiology and the Quest for Human Nature (MIT Press, 1985).

Wishing on a star

Andrew H. Knoll

Nemesis: The Death-Star and Other Theories of Mass Extinction. By Donald Goldsmith. Walker, New York: 1985. Pp. 166. \$15.95.

A HERD of hadrosaurs grazes calmly, apparently unconcerned by a nearby tyrannosaur. Suddenly, a fireball streaks across the sky; there is a huge explosion, a conflagration erupts and a curtain of darkness descends across the landscape. In the six-month night that follows, the dinosaurs all die, as do innumerable, less evocative creatures. Science fiction? Some scientists think so, but an impressive body of data supports the idea that this is how the Cretaceous Period ended. More controversial evidence suggests that similar events have occurred every 26 million years or so, extirpating much of the Earth's biota and realigning patterns of evolution.

Donald Goldsmith has prepared a layman's guide to the intricacies of the extraordinary and controversial Nemesis theory of periodic mass extinctions. He begins appropriately with an exposition of the Alvarez hypothesis that a bolide hit Earth and precipitated mass extinctions 65 million years ago, then proceeds to the case for periodicity in the fossil record of extinctions, as well as periodicity arguments for impact craters. Dissenting views are presented, but sceptics come off as wrong-headed curmudgeons who "snort in disapproval" (really) at statistical arguments. This may reflect measured scientific judgement, but Goldsmith's treatment is too brief for readers to draw their own conclusions.

From acceptance of the principle of periodicity, it is but a short jump to cosmic explanations for mass extinction, and Goldsmith presents informative discussions of both the Nemesis theory in which

cometary orbits are perturbed by an as-yet-undetected companion of our Sun and the competing idea of cometary perturbations caused by the Solar System's regular passage through the galactic plane. Goldsmith takes the galactic oscillation hypothesis more seriously than most astrophysicists of my acquaintance, but his heart clearly lies with Nemesis. His precise description of the "death star" is fascinating — oddly reminiscent of St Augustine's vision of the City of God, but with physics lending rigour to faith.

Goldsmith's perspective is decidedly that of an astronomer. He is at his best explaining how fair-minded scientists using the same imperfect data can arrive at radically different estimates of periodicity. Less satisfactory is his discussion of the data sets themselves — their problems, limitations and biological implications. Indeed, in *Nemesis* mass extinctions are largely an interesting means of learning something new about the cosmos. Certainly, they are potentially that, but they are much more.

Goldsmith's leitmotiv is that the personal (and often unscientific) motivations of individual scientists collectively result in the system of observations, experiment and debate that is science. Several of his points are well made, but because he discusses this subject in abstract terms rather than through the personalities of the protagonists in the debate about Nemesis, his lessons have a detached, pedagogical ring to them. Perhaps this is why the book is ultimately less riveting than the best scientific chronicles, such as *The Eighth Day of Creation* or *Ice Ages: Solving the Mystery*. Still, *Nemesis* provides non-specialists with a concise introduction to one of this century's most exciting and potentially far-reaching debates about Earth history. □

Andrew H. Knoll is a Professor in the Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Shadowgraphs of a mixing layer from flow visualization experiments in fluid mechanics. The illustration is reproduced from Turbulence and Random Processes in Fluid Mechanics, by M.T. Landahl and E. Mollo-Christensen, published by Cambridge University Press. Price is £20, \$34.50.

Different sides of the sexes

Marian Dawkins

The Female Animal. By Irene Elia. Oxford University Press: 1986. Pp. 319. £19.50.

Females of the Species. By Bettyann Kevles. Harvard University Press: 1986. Pp. 254. \$20. To be published in Britain next month, £16.95.

THERE are some women who believe that biology places too much emphasis on the importance of male behaviour and male interests, brought about because so many biologists are men. Two such women who consider that the time has come to redress this imbalance and put the female point of view have independently written books which attempt to do just that. They cover much the same ground, and make many of the same points, but they do so in very different tones.

In *The Female Animal*, Irene Elia argues that the female role in reproduction, which has sometimes been seen as exploitation of females by males, should more properly be seen as female control. She argues that females and their nurturing, mothering qualities have dominated evolution and given rise to all the major evolutionary advances since the beginning of life on Earth. Females give their eggs food and shelter. By providing them with shells or nests, or ultimately retaining them in their own bodies, females enabled living forms to come out of the water. It was females, not males, that conquered dry land. And it was the care that females gave to their helpless young that eventually gave rise to the social communities of birds and mammals. The advantages of mothering and the demands of care and protection of the young led to the evolution of increasingly complex nervous systems, better senses to detect danger that might threaten the young and behaviour to feed and keep them warm. Eventually, argues Elia, the intelligence and cooperation which female mammals evolved for the extended mothering of their young led to all we regard as most worthy in human behaviour, unselfishness and moral altruism for example. Males are portrayed as the secondary sex, either as mere reproductive appendages like the male Holboell anglerfish, or fighters over

valuable females or, at best, as following the female lead and caring for young themselves.

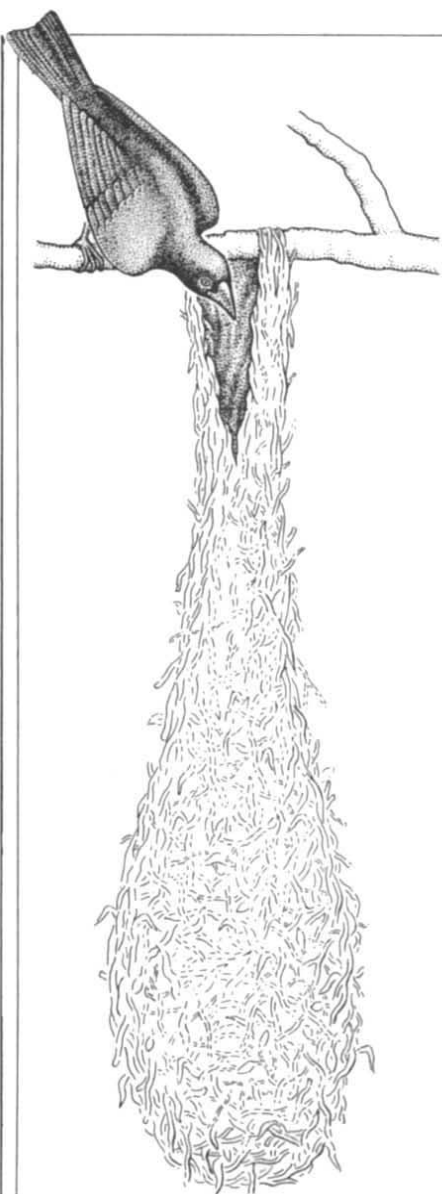
Elia makes her case forcefully, although many of her points, despite what she appears to think, are quite widely accepted. Few biologists would dispute the importance of maternal care or of female reproductive processes such as vivipary. Many field studies have now demonstrated the importance of female behaviour, including female competitiveness, in social structure. But Elia obviously feels that, to use her own words, the tale should be retold and that the call of "females first" needs to be reiterated more loudly and more clearly than it has been before.

She does this by documenting, group by group, the reproductive strategies to be found in various animals, with emphasis throughout on the female role. She then argues that, in humans, the domestication of plants and animals led to a crisis in the relation between men and women; what she calls the world's most mothering female became dominated by men who possessed the physical strength to secure by force both women and the fruits of the new agriculture. The book ends with these words:

since we became domesticators of plants and animals, we have been constrained to accede to and sometimes contribute to the use of force, surveillance, infanticide, marriage, incarceration, infibulation, and social-psychological conditioning to ensure that the amniote-gestator-lactator legacy of female mammals be used at male behest.

This strident style detracts rather than adds to her case. And is marriage really to be seen as being to male advantage only? Is infanticide really imposed on females by males?

Using many of the same examples, but writing in a much more sober way, Bettyann Kevles also looks at evolution and courtship from the females' point of view. She, too, takes male biologists to task for their male-centred view. The exception, for her, is Charles Darwin, who Kevles praises for emphasizing the role of female choice. She describes the reproductive strategies of various species, the great variety of parental care and the even greater variety of roles that females adopt throughout the animal kingdom. She writes well, though most of her examples are to be found in standard textbooks on the subject. Like Elia, she is concerned to "tell the tale" again, through female eyes, but her treatment is altogether more balanced and she makes her points without antagonizing the reader. She also discusses what she calls "the other side of the coin" — the way in which females dominate other females and sometimes even kill their offspring. Females emerge from her book as just self-centred and manipulative of other individuals as males



Devious female — a cowbird using the nest of the oropendola for her own eggs.

From *Females of the Species*

are. Females, for her, are equal players in the evolutionary game, neither controlling or controlled by males.

This view of an evolutionary balance between the sexes is far more convincing than Elia's picture of control and dominance by one or the other, but I have to confess to having had some difficulty in reviewing both of these books. Biology does not seem to be nearly as male-centred as Elia and Kevles think it is. Not just Darwin, but Fisher and more recently Lande, Zahavi and Hamilton, amongst others, have placed a great deal of stress on the importance of female choice and female behaviour, sometimes on very little evidence. If anything, it seems to me, female choice may turn out in the end to have been given too much prominence by biologists, rather than too little. □

Marian Stamp Dawkins is Fellow in Biological Sciences at Somerville College, University of Oxford, Oxford OX2 6HD, UK.

• *Animal Behavior and its Applications*, by Derek V. Ellis, is an account of "the daily activities of animals . . . and how we as humans impact [sic] and manipulate those activities". The book is published by Lewis Publishers, 121 South Main Street, Chelsea, Michigan 48118, USA, and is distributed in Britain and the rest of Europe by John Wiley, Chichester, UK. Price is \$29.95, £27.55.



NEW SEASON TITLES

LOOKING AT VERTEBRATES

A Practical Guide To
Vertebrate Adaptations
Elizabeth Rogers

Cased Illustrated £9.95 net
ISBN 0 582 45086 1 March 1986

UNRAVELLING ANIMAL BEHAVIOUR

Marian Stamp Dawkins

Paper Illustrated £6.50 net
ISBN 0 582 44691 0 March 1986

EVOLUTION AND CLASSIFICATION

The Reformation Of Cladism
Mark Ridley

Paper Illustrated £6.95 net
ISBN 0 582 44497 7 April 1986

INORGANIC CHEMISTRY Second Edition

A G Sharpe

Paper & Cased Illustrated
£12.95 & £25.00 net
ISBN 0 582 44688 0 & 0 582 44689 9
April 1986

KAYE & LABY'S TABLES OF PHYSICAL AND CHEMICAL CONSTANTS

Fifteenth Edition

Cased Illustrated
Probably £17.50 net
ISBN 0 582 46354 8 April 1986



For full details of these and related titles,
please write to Clive Weatherley, Longman
Group Ltd., Longman House, Burnt Mill,
Harlow, Essex, CM20 2JE, UK.



Ordered thoughts on taxonomy

Michael T. Ghiselin

Evolution and Classification: The Reformation of Cladism. By Mark Ridley. Longman: 1986. Pp. 201. Pbk £6.95, \$16.95. From 1 May Longman science books will be distributed in the United States by John Wiley, New York.

BIOLOGY is a branch of strictly objective natural science — at least it ought to be. But in this book Mark Ridley argues at great length that the leading schools of taxonomy have failed to live up to this ideal standard, with the noteworthy exception of "cladism" as it was originally formulated by its founder, Willi Hennig. The others, namely numerical phenetics, the evolutionary systematics of Simpson and Mayr, and the "transformed" version of cladism, are sadly deficient in this respect.

The differences between the various schools are philosophical, and Ridley makes it clear that bad philosophy is the culprit for all that is wrong in taxonomical thinking. Most of the points he raises are by no means new, but the detailed examination of the issues provides a compelling argument for objectivity in classification. For a system to be objective it needs what Ridley calls a "principle of justification", such that we classify the objects themselves not the subjects ourselves (for example, a ranking of ink blots in the order of their obscenity obviously projects the properties of the subject into the system). Less obvious is the point that the concept of "similarity" is open to the same basic criticism; there is no way to decide which objects are "more similar" upon anything but a subjective basis. The failure of numerical pheneticists to come up with an objective criterion for what they called "overall similarity" indicates that their fundamental premise was in error.

The periodic table of the elements is justified as an objective system by the properties of atoms and the laws that govern them. Genealogical classification is justified by the fact that the taxonomic groups stand to each other as parts of a genealogical nexus. Ridley examines other possible principles of justification, such as Ideas in the Divine Intellect, and finds them unsatisfactory. Evolutionary systematics he considers imperfect, because it is tainted with the same kind of subjectivity that affects phenetics. This leaves cladism, but not the "transformed" version, which "has been greedily incorporated into that great media event, the crisis of Darwin" (p.10).

Ridley is not the first to advocate *strictly*

genealogical classification. Darwin did. Unfortunately, biologists have persisted with the traditional practice of obscuring relationships by removing modified organisms from the groups that gave rise to them. A mixture of genealogy and "amount of difference" does not provide the kind of arrangement which a scientist needs for doing research, because the two criteria lead to ambiguity. Consider, for example, our own genus, the closest relatives of which are chimpanzees and *Gorilla*. Putting *Homo* in one family and what might be called "the rest of us apes" in another, gives the false impression that our lineage branched off before the common ancestor of gibbons and chimpanzees. Those of us who do research on the history of life have not been unaware of this problem, and have used phylogenetic trees instead of formal classification systems wherever possible. A strictly genealogical system has certain disadvantages with respect to nomenclature, which Ridley perhaps should have discussed, but the price he has paid is not unreasonable. As a professional student of evolutionary mechanisms, he needs to know the facts about evolutionary history, not some taxonomist's evaluation of them.

Transformed cladism, which attempts to do without evolutionary justification, is Ridley's prime target. The transformed cladists' philosophy is a throwback to pre-Darwinian idealistic morphology, combined with a naive inductionism that rejects the modern hypothetico-deductive method out of hand. According to this ideal of science, we are supposed to wait until the "patterns" have been worked out before we begin to study "process", which is like saying that astronomers should have waited until the epicycles were known in exhaustive detail before investigating the laws that govern celestial bodies.

Ridley maintains that the transformed cladists have argued their case on the basis of sophistry, bad metaphysics and outright misrepresentation; indeed, so successful has been their conspiracy of silence with respect to the use of functional anatomy that Ridley himself seems unaware of the extensive literature on that topic, and of its routine application by the best workers for more than a century. The use of vestigial structures to indicate the direction of evolutionary change is but one obvious example — the teeth in the jaws of baleen whale embryos are obviously on the way out. It is not out of character that those who ignore legitimate scientific data when doing research go so far as to pretend that such data do not even exist when writing polemics. But this neglect is not characteristic of transformed cladists alone. Classifying on the basis of "pure morphology", which leaves out all consideration of function, is symptomatic of the pervasive

influence of typological thinking.

All the leading schools of taxonomy have taken it for granted that classification is somehow based upon putting similar things together on the basis of shared attributes, and nothing else. There is little hope for rectifying the situation until the falsehood of this premise is generally recognized. Classifications ought to be based upon a scientific evaluation of any data that happen to be germane; that is, upon scientific knowledge as a whole, drawing upon physics and chemistry, as well as the laws, principles and historical particulars of biology and geology. It means thinking like a historian, asking what has happened and why, and formulating hypotheses and gathering data to test them.

There are innumerable opportunities available for doing the job well. At the moment the study of large molecules apt to evolve divergently is beginning to yield very impressive results. Charles Sibley's monumental work on DNA $\times$ DNA hybridization in birds has at last provided us with a rigorous phylogeny of the group, this in spite of the fact that it does not

employ shared derived characters and has therefore been spurned by the cladists. It works instead because, as time passes, homologues cease to be identical, and the relevant parameters can be measured. Likewise Carl Woese's frontal attack on prokaryote ribosomal RNA has demolished the claims of pheneticists that bacterial phylogenetics is simply unknowable. Norman Pace has developed a technique whereby hundreds of nucleotides of this same material can be read off a single gel, and data on the phyla and classes of Metazoa are rapidly accumulating in his laboratory. Once the basic phylogenetic tree has been worked out, biologists will then find it much easier to place the whole range of their findings in an evolutionary context. When that happens, the value of genealogy will be so clear that its alternatives will no longer be taken seriously. Good science generally wins out over bad philosophy, but it takes a long time. $\square$

Michael T. Ghiselin is MacArthur Fellow in the Department of Invertebrate Zoology, California Academy of Sciences, Golden Gate Park, San Francisco, California 94118, USA.

Dancing in time

Leonard Krishtalka

The Great American Biotic Interchange. Edited by Francis G. Stehli and S. David Webb. Plenum:1985. Pp.532. \$75, £61.64.

FOR THE past 90 million years (Myr) North and South America have danced a biological and geological *pas de deux* — long periods of drifting in isolation punctuated by brief encounters, animal invasions and faunal interchange. Until quite recently, much of the palaeobiological ballet of the Americas was scattered throughout a vast literature. Not any more. *The Great American Biotic Interchange* is an outstanding reference work on 90 Myr of New World faunas, geochronology and palaeogeography, especially of South America. It is appropriately dedicated to George Gaylord Simpson and complements his more popular book on the subject, *Splendid Isolation: The Curious History of South American Mammals*, published by Yale University Press in 1980.

The volume opens in the late Cretaceous. By that time plate tectonics and a spreading Atlantic sea floor had torn the Americas from Europe and Africa and waltzed them into momentary contact across a Caribbean island arc. The fossil record hints that South American didelphid marsupials hopped into North America for the first time, while northern lizards, snakes, duckbilled dinosaurs and a primitive placental ungulate entered South America.

The Americas jeté apart for the next 87 million years, their faunas evolving in near isolation from one another. North America became dominated by placental mammals: primates (initially), insectivores, rodents, rabbits, perissodactyls, artiodactyls, mastodons and a legion of now extinct orders. South America hosted the explosive radiation of a horde of marsupials, edentates (sloths, anteaters, armadillos) and unusual, now extinct, placental ungulates (hoofed herbivores: didolodons, astrapotheres, litopterns, notoungulates and pyrotheres).

Three events pierced the calm of faunal isolation. The first, of least consequence, seemingly occurred in the mid-Palaeocene (58–61 Myr ago), when edentates and notoungulates appear to have island-hopped to North America from the south, left a few tantalizing bony remains and continued on to Asia. The second episode, during the mid- to late Oligocene (33–28 Myr ago), involved the sudden appearance in South America of platyrrhine primates and caviomorph rodents, immigrants on a flotilla of "Noah's Arks" — biotic rafts — from North America or Africa. The third event, during the late Miocene (7–9 Myr ago), was a portent of things to come: a filter bridge between the Americas allowed some racoons and rodents to pass south and two groups of sloths to move north.

Three million years ago the Panamanian land corridor surfaced and all pretence of faunal isolation ended; the Great American Interchange was underway. Among the first waves of northern mammals to invade South America were

canids, ursids, mustelids, felids, tapirs, equids, camelids, cervids and mastodons. In turn, rodents (porcupine, capybara), edentates (glyptodonts, ground sloths, armadillos), a llama, an opossum and a flightless bird dispersed into North America.

Peppering this history are palaeoecological questions: the convergent suites of morphological adaptations among mammalian ecological vicars in North and South America; the allegedly precocious evolution in South America of savannahs and grazing ungulates with high-crowned dentitions; the effects, if any, of the northern placental invaders on southern faunas; and the ancestors, provenance and date of arrival of the first South American primates and rodents.

Unlike the scattershot approach of most symposium volumes, the material in the book is organized in textbook fashion (19 chapters in five sections). Sections I and II examine Caribbean plate motions; plate evolution; evidence of inter-American mammals in the Palaeocene; herpetofaunas from the Cretaceous and the Cenozoic; and geochronology, mammalian biochronology and interchange events for the same periods. The last two (by R. Estes and A. Báez, and L.G. Marshall) are excellent and exhaustive treatments.

Section III covers the Tertiary evolution and dispersal of North American mammals; the origin and diversification of South American mammals, ungulates, and primates and rodents. Here nationalism seems to reign. Most American palaeontologists, including A.E. Wood who deals with primates and rodents, claim a New World mid-Tertiary origin for the South American platyrrhines and caviomorphs. Old World palaeobiologists favour an African origin.

The last two sections are the climax — close looks at the physical and biotic dynamics of the Great American Interchange over the past 3 Myr: Caribbean tectonics; the Central American land bridge, climate and sea level; the inter-American dispersal of invertebrates, plants and their bee pollinators, fish, amphibians and reptiles, birds, all mammals, and desert and grassland mammals.

Accompanying the integrated accounts of palaeobiology, biogeography, ecology, systematics and geology is a staggering amount of information, such as the detailed compendia of fossil vertebrate occurrences in South America. *The Great American Biotic Interchange* is a comprehensive, up-to-date and invaluable resource for any student interested in the past and present biotas of the Western hemisphere. $\square$

Leonard Krishtalka is Associate Curator in the Section of Vertebrate Fossils, Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, Pennsylvania 15213, USA.

Eastern adventures revisited

Eric Ashby

The Head-Hunters of Borneo. By Carl Bock. *Oxford University Press: 1986.* Pp.344. Pbk £6.95, \$9.95.*

A Naturalist in Borneo. By Robert W. Shelford. *Oxford University Press: 1986.* Pp.331. Pbk £4.95, \$8.95.*

The Field-Book of a Jungle-Wallah. By Charles Hose. *Oxford University Press: 1986.* Pp.216. Pbk £4.95, \$8.95.*

THE island of Borneo, three times the size of Great Britain and larger than the State of Texas, still evokes an aura of mystery and chilly romance. It was the home of head-hunters and cannibals. Part of it, Sarawak, was annexed, at the invitation of the native ruler, by an adventurer brought up in the English city of Norwich, who took over its government in 1841 and founded a dynasty of rajahs named Brooke. The Brookes were shining examples of the process the present British government call "privatization": they suppressed piracy and other crime, exacted a poll tax from the indigenous inhabitants, and promoted trade in rattan, gutta-percha, pepper and camphor. Behind coastlines fringed with mangroves, and deltas hidden behind nipah palm and casuarina, there were virgin tropical forests and tribes untainted by Western civilization.

The aura of mystery is being rapidly dispersed. Any watcher of nature-films on television can now know more about the wildlife, human and otherwise, of Borneo than all but a few hundred Europeans and Americans 50 years ago. So what have we to learn from the records of naturalists and travellers who were not equipped with movie cameras, zoom lenses, tape recorders and helicopters? Evidently something, or Oxford University Press would not have re-issued these three books.

The main thing to be learnt is that many of the marvellous intricacies of animal behaviour and the sophistication of social life among so-called savages have been known to naturalists and anthropologists for the greater part of a century, observed by naked eye, through a hand lens or binoculars, and recorded in print and not on film. This is not to belittle the achievements of the present generation of naturalists, who have interpreted these discoveries to the public and who have, of course, greatly extended them; but it does deepen one's respect for the pioneers who went into the jungles of Borneo without the benefit of anti-malaria drugs (except

quinine), or prophylactics against dysentery, or radio communication to summon help in danger, or any recording device except a heavy plate camera and a sketch-book. These books were written by three very different people. In comparing them, the main interest is not in the facts they record; rather, it is in the impression that Borneo made upon the writers.

Carl Bock's book, published in 1881, is the oldest. It records two expeditions, one made in 1878 to Sumatra to collect birds for the Marquess of Tweeddale, and the other in 1880, commissioned by the Governor-General of the Dutch East Indies, to travel through the Koetei region of south-east Borneo, at that time a Dutch colony. It is the best known of the three books (partly because of its sensational title), but as a record of natural history it is the least interesting. Bock was essentially a journalist and travelled in order to write about his travels, not to make discoveries. He was much more interested in the people than in the animals and birds and insects, and his interest in the people was largely superficial: what did they wear, how did they hunt, what happened when they held festivals and funerals? He had a journalist's taste for the bizarre and unusual; his most persistent search was for the legendary *orang buntut*, Darwin's "missing link" — a tribe of men with tails.

Bock wanted to stay among head-hunters and see samples of their preserved heads; in this he was not disappointed. The notorious Dyaks were hospitable to him (he was accompanied by a retinue of 52 "tame" Dyaks and 22 armed soldiers with the authority of the Sultan of Koetei to back him). He was present at the Sultan's birthday party and describes it with zest: the silver tea service, and the eau de Cologne for the ladies (four wives and 38

concubines). Although the Sultan was a Muslim and drank only soft drinks, a chest of champagne was produced at Bock's suggestion, as a fitting beverage for the occasion.

The book is a racy read and it brings one close to a remarkable man. He was only 29 years old, quite fearless, taking great risks without boasting about them, and although he couldn't speak a word of the language he was able to convey a sense of friendliness towards the natives that won their confidence. He disliked the dirt and squalor, the food and the mosquitos, but he recorded all these miseries cheerfully and objectively. As a naturalist, however, he was third rate, his observations mostly trivial, and his achievement slight. As a traveller's tale the book is well worth reading; but Bock was not the first explorer of this region of Borneo, nor is his book a perceptive record of its wildlife or its inhabitants.

Robert Shelford's *A Naturalist in Borneo* was first published in 1916, a generation after Bock had visited Borneo. Shelford was a biologist from Cambridge who set out upon an academic career in 1895. Two years later he interrupted his studies to become Curator of the Sarawak Museum (another initiative of the Rajah Brooke), and stayed there for seven years. It was a wonderful opportunity for a naturalist. The task he set himself was not merely to collect specimens for a museum, it was to study creatures as he found them in the forests and swamps. He was primarily an entomologist, but at the end of his contract he brought back to England an immense amount of material on the behaviour of animals and birds, observations on mimicry in insects (at that time a controversial topic) and some penetrating studies of the social life of the indigenous people. For the rest of his short life — he



Home for the dead — Bornean mausoleum, depicted in a lithograph (coloured in the original) based on a drawing by Carl Bock. (Reproduced from the section of plates in *The Head-Hunters of Borneo*.)

* To be published in the United States later this year.

died in 1912 at the age of 40 — Shelford worked in Oxford, painstakingly sorting out his notes and exquisite drawings and photographs. He died before the work was ready for publication. Edward Poulton, then professor of zoology at Oxford, edited the work and this is the volume which has now been re-issued. It is a tidy and systematic description of the mammals, birds, snakes, cockroaches (a special interest he had), beetles and ants in Sarawak; a thoroughly professional job, but told in an engaging way which will appeal to the same wide audience that finds David Attenborough's films so compelling.

If I had to recommend only one of these books, it would be the third, by Charles Hose. Despite its silly title, which leads one to expect pompous recollections from some bewhiskered ex-Indian civil servant from Cheltenham, it is learned, charming and scientifically credible. Of the three, it is the only one that gives the impression that here is a man permeated by the environment of Sarawak. This is not surprising, for Charles Hose was a senior official in Rajah Brooke's government and a member of the Supreme Council. He served there from 1884 to 1907, although his notebooks were not published until 1929. His style is conversational, as if conducting a leisurely tour: sailing along the Bornean coast he draws the reader's attention to fish, to turtles, to birds, with acute observations about each of them, clearly the fruit of years of experience. His comments about the people, too, are of a much higher quality than those of Bock or even Shelford; Hose spoke their languages, knew them as individuals and wrote about them with affection, as fellow citizens, not as subjects of the "white man's burden".

This book, I believe, reproduces something for which there are few modern equivalents, namely the deep-rooted intimacy of an outsider with an exotic foreign country. Even the most lavishly equipped film-team cannot acquire that intimacy; and these days fewer and fewer Westerners have opportunities to spend decades in such alien societies. The surprising feature of *The Field-Book of a Jungle-Wallah* is that it is as authoritative and precise in its observations of nature as Shelford's book, although natural history for Hose was a hobby, to be enjoyed in the time he could spare from administration and the duties of high office.

All of these books are, in their different ways, most enjoyable reading. They will also remind the maker of nature-films that he owes a debt to his forerunners who put laboriously into print what they, more swiftly, now put on to celluloid. □

Lord Ashby, formerly Professor of Botany at the University of Sydney and the University of Manchester, is a Fellow of Clare College, Cambridge CB2 1TL, UK.

Colour impressions of history

J.D. Mollon

Jacob Christoph Le Blon 1667–1741: Inventor of Three- and Four-Colour Printing. By Otto M. Lilien. *Anton Hiersemann, Postfach 723, Rosenbergstrasse 113, D-7000 Stuttgart 1, FRG: 1985. DM 180.*

DESPITE the seemingly endless variety of our visual sensations, our colour vision depends on the presence in the retina of just three classes of cone, each containing a different photo-sensitive pigment. It is for this reason that colour reproduction is relatively easy: in each local area of the image, we must merely arrange that three printing inks or three television phosphors stimulate the viewer's cones in the same ratios as would the original object.

The principle of three-colour reproduction was exploited in practice before its theoretical basis was understood. The three-receptor theory of vision did not emerge until the end of the eighteenth century, but colour printing had already been brought almost to perfection by its quixotic inventor, Jacob Christoph Le Blon, a miniaturist and mezzotint engraver who was born in Frankfurt in 1667. Horace Walpole said of Le Blon:

He was... of surprising vivacity and volubility, and with a head admirably mechanic, but an universal projector, and with at least one of the qualities that attend that vocation, either a dupe or a cheat;... perhaps like most enthusiasts he was both one and t'other.

Le Blon has long been a shadowy and enigmatic figure and we may warmly welcome O.M. Lilien's careful biography.

Lilien has eschewed the often fanciful accounts of Le Blon in the secondary literature, and, combining scholarship with enterprise, has scoured Europe for contemporary material, in civil archives and in printed sources. He is thus able to provide us with a firm, though somewhat spare, reconstruction of Le Blon's career. Many of the more important documents are reproduced. Lilien discusses in detail the editions of Le Blon's book *Il Coloritto*, in which the principle of trichromatic colour-mixing is set out. He reprints in full the text of the first edition. However, although Lilien is a historian of printing, there are some gaps in his account of Le Blon's printing technique. It took Le Blon a long time to find inks with the necessary properties, but Lilien does not discuss their composition, although they are identified in eighteenth-century sources (for example Dossie's *Handmaiden to the Arts* of 1764) and although they could be identified spectrophotometrically from surviving prints.

One of the most important documents considered by Lilien is a holograph prospectus for the "Picture Office", the ill-fated printing company that Le Blon established in London. In translating from this French manuscript to English, Lilien corrupts one very interesting passage (cf. pp.28, 30 and 115 of his book). We can readily forgive him, since neither French nor English is his native language. But the passage throws an intriguing light on Le Blon's understanding of trichromacy. My own translation would be as follows:

In searching for general rules that would allow one to reduce to solid, reliable and easy practice that part of painting called *Coloris* by painters, I perceived that my project would be prac-

Of Preliminaries.

COLORITTO, or the *Harmony of Colouring*, is the *Art of Mixing COLOURS*, in order to represent naturally, in all Degrees of painted Light and Shade, the same FLESH, or the Colour of any other Object, that is represented in the true or pure Light.

PAINTING can represent all visible Objects with three Colours, *Tellow, Red, and Blue*; for all other Colours can be compos'd of these Three, which I call *Primitive*; for Example,

<i>Tellow</i>	}	make an <i>Orange Colour</i> .
<i>Red</i>		
<i>Red</i>	}	make a <i>Purple and Violet Colour</i> .
<i>Blue</i>		
<i>Blue</i>	}	make a <i>Green Colour</i> .
<i>Tellow</i>		

And a *Mixture* of those Three Original Colours makes a *Black*, and all other Colours whatsoever; as I have demonstrated by my Invention of *Printing Pictures and Figures with their natural Colours*.

I am only speaking of *Material Colours*, or those used by *Painters*; for a *Mixture* of all the primitive *impalpable Colours*, that cannot be felt, will not produce *Black*, but the very Contrary, *White*; as the Great Sir ISAAC NEWTON has demonstrated in his Opticks.

White, is a Concentring, or an Excess of Lights.
Black, is a deep Hiding, or Privation of Lights.

Mixed interpretation — the opening page of Le Blon's *Il Coloritto*. The book was first published in 1725.

ticable if I could find among material colours the perfect primitive colours. Having then at last found material colours that came so close to the primitive colours that there was no tint that could escape me and not be reproduced by my colours, I began to realize that, according to these principles, painting could perfectly represent all visible objects not only by the brush but also by the printing press.

There is a strong hint here that by 1721 Le Blon had already distinguished in his mind between the most saturated material colours and hypothetical primitives or primaries. This important distinction was to become critical in the later understanding of colour vision and it took Clerk Maxwell and Helmholtz to grasp that each of the hypothetical primaries would represent unique excitation of one receptor type — something that a physical stimulus can only approach.

There is still much work to be done on Le Blon, on his techniques and on his theoretical understanding. But Lilien has provided us with an authoritative and richly illustrated biography that will not quickly be surpassed. □

J.D. Mollon is a Lecturer in the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

Sold on application in France

Robert Fox

From Knowledge to Power: The Rise of the Science Empire in France 1860–1939. By Harry W. Paul. Cambridge University Press: 1986. Pp. 415. £32.50, \$49.50.

THE most casual visitor to France cannot fail to observe the signs of modernity that spring from a confidence in science and technology. A decisive national policy for science, implemented since the Second World War through a series of five-year plans, has borne fruit ranging from the spectacular new Cité des Sciences et de l'Industrie at the Parc de la Villette to such conveniences of everyday life as an incomparably efficient telecommunications system and a high-speed train that covers the 300 miles between Paris and Lyons in two hours. Clearly, in today's France science is at the centre of public and private life, and there is every sign (not least in the overwhelming popularity of the scientific options in the *baccalauréat*) that what has happened is irreversible.

For well over a decade, Harry Paul has explored the early history of France's emergence as a pace-maker in European science and technology. His writing has always been richly documented and witty, and this book is no exception. It elucidates in unprecedented detail the process by which, between the 1860s and the beginning of the Second World War, the French scientific community grew not only in size but also in the funds it consumed and in the attention it commanded in the fashioning of governmental and industrial policy. There were, of course, other economically advanced countries where comparable growth occurred. But what happened in France is made especially remarkable by the circumstances; the period treated by Paul was punctuated by financial crises, crippling bouts of political conflict and wars, in 1870 and 1914–1918, which in different though equally terrible ways sapped national morale.

Paul recognizes, at least by implication, the complexity of the historical problems that lie in this paradox. For him, the heart of the matter lies in a fundamental transformation in the role of the *savant*. From being pre-eminently men of knowledge, French scientists became men of power. Marcellin Berthelot, Paul Bert and Adolphe Wurtz were among those who secured power by political means, serving either as ministers or as prominent members of the Senate or the Chamber of Deputies. But even greater power was achieved through the repeated and eventually successful peddling of the notion that science was, above all, useful.

The main articulators and beneficiaries of this approach were the professors in the faculties of science, who from the later nineteenth century began to develop a degree of involvement in applied studies that would have been unthinkable in the original conception of the Napoleonic University. Between 1890 and 1914, more than 20 institutes of applied science were created either within or as annexes to the science faculties. Most of the new foundations were successful, whether we judge them by the numbers of students who enrolled or by the financial support they elicited from industrial and agricultural interests and from councils intent on promoting the local economy and civic pride. At Nancy, for example, more than three-quarters of the students in the faculty of science on the eve of the First World War were working in one or other of the specialized institutes (for chemistry, electrical technology, aeronautics and brewing), and a high proportion of the faculty's income came in the form of grants from the departments of the region and from industrial benefactors, among them Ernest Solvay.

Generally, Paul takes a benign view of this shift in the balance between the pure and the applied, and there are important respects in which he is plainly correct. How else could the faculties have justified the new buildings and new posts that were provided in such abundance about the turn of the century? And how else could they have suddenly begun to attract students after decades in which they had done little more than mount polite lectures for a bourgeois public? Yet the losses and the failures deserve attention (notably in Britain, where the universities are being asked to strike out on a road not unlike the one that the French faculties took a century ago). There is some evidence, for example, that the emphasis on utility in higher education weakened France's position in fundamental research and may not even have satisfied the expectations of industrialists and agriculturalists. On the latter point, Paul's discussion of the employers' lack of enthusiasm for the men and ideas emerging from the academic system between 1919 and 1939 is tantalizingly brief though it is no less arresting as a cautionary tale. The lesson, it seems, is that educational reform is an eternally alluring panacea but that, in itself, it can do little (except perhaps in the very short term) to solve the problems of a flagging economy.

By the 1930s, therefore, French academic science bore the scars of at least partial failure. But, by then, the growth of the scientific community, combined with the public acclaim of such great figures as Marie Curie and Frédéric Joliot and the spectre of economic depression and war, had created a receptive ear in ministerial circles. The scientific lobby duly trium-

phed, first in the enhanced funding for research that came with the Popular Front and then, in 1939, in the creation of the Centre National de la Recherche Scientifique.

Throughout the book, Paul balances his dominant expansionist theme with reminders that the rise of what he calls the science empire was neither natural nor inevitable. His discussion of the reactions, within and outside the scientific community, to the potentially disquieting bond between science and positivism reminds us, for example, that it was as important in the nineteenth century as it is now for *savants* to maintain a wholesome public face. In all but the most conservative religious circles, it seems, the ideological neutrality of science came to be accepted, despite the potent mix of Darwinism and materialism that was purveyed in much biological writing in the later nineteenth century. Once doctrinal fears were allayed, the Catholic universities did not hesitate to embrace science; indeed, they turned to the task of training men for the modern industrial world quite as enthusiastically as any of the aggressively secular faculties of the state.

Paul also underlines the importance of war and other misfortunes in providing new opportunities for scientists to convince the public and governments of the capacity of science to serve the national interest. The First World War, in particular, seems to have been the occasion for the deployment of every conceivable rhetorical device in support of science. For the scientific community, the rewards were to be seen in the favourable attention of ministers, sponsored research and an opportunity of appearing before the public in an exhilaratingly heroic guise. (The question whether the mobilization of *savants* really did help the war effort is, of course, a quite separate one, and it is largely ignored in this book.)

With the publication of *From Knowledge to Power*, the study of French science in a period which until recently was neglected in favour of the more glamorous episodes of the seventeenth century and the Revolution and First Empire has been notably advanced. Harry Paul has always been regarded as an adventurous historian, and now, once again, he has worked barely tilled soil to produce a book of characteristically absorbing interest. □

Robert Fox is British Academy Reader in the Humanities, Department of History, University of Lancaster, Lancaster LA1 4YG, UK.

● Shortly to be published by University of California Press is *Science in the Provinces*, by Mary Jo Nye. The book concentrates on university science in France during the period 1860–1930, the author arguing that “provincial scientists and administrators often took the lead in developing research and organizational initiatives”. Price will be \$39.95, £33.95.

Bits and pieces of science policy

Robert Walgate

The European Scientific Community. By Ros Herman. Longman: 1986. Pp.201. £18.50. Longman appear not to know whether this book is available in the United States.

FROM 1550 to 1850, science was a European affair. Now, more or less, it is American. Why? Ros Herman, the senior writer on science policy for *New Scientist*, immediately places a simple, nagging question in the reader's mind with the first words of her book; but the question remains unanswered, though it hovers over every page.

Not that Ms Herman explicitly attempts to answer her question. Her book, as she writes later, aims to provide a backdrop, a catalogue and interpretation of the myriad national and international institutions which are involved in the government of European science. She looks at budgets, manpower, policy-making structures and policies throughout Europe, and sets them in a historical perspective. In the end, she succeeds in charting the maze of faint crossing and parallel tracks along which European science lost its pre-eminence.

Of course, European science is not dead; its decline is only relative, and there are isolated areas where Europe continues to lead (usually because of extraordinary drive on the part of an individual scientist). However, we need to explain why European papers now make up only a small part of the research published in leading journals. On p.13 of her book, Ms Herman quotes Vannevar Bush, speaking in the United States in 1945:

We can no longer count on a ravaged Europe as a source of fundamental knowledge. In the past we have devoted much of our best efforts to the application of such knowledge which has been discovered abroad. In the future we must pay increased attention to discovering this knowledge for ourselves.

Is the resolution implicit in this remark what led to American greatness in science, and relative European decline? (And, if so, what about the Japanese, who say much the same thing today?) It was not necessary for European science to decline *per se*: the great explosion of American science simply, bit by bit, began to shift the glory, the important discoveries, the prizes across the Atlantic.

There were — and are — special circumstances in the United States. By 1945 the American scientific community had been swelled by the wartime flood of European émigrés (not long ago, for example, it was possible to trace all the lead-

ing high-energy physics groups in the United States back to Fermi and his pupils). The open, meritocratic and competitive society encouraged science of the same character. With its own firm science base, leavened by the wisdom of the émigrés, with the determination of people like Mr Bush (and the money that followed) and with a mobile society of a couple of hundred million souls, America was set fair.

The European "answer", after a savage nationalistic war, was to set up multitudinous, national science planning and policy bodies, scattered over the continent (and Britain). Everyone knew that "science had won the war", that science was important. The trouble was that Europe was — and remains — in comparatively small pieces, with a British science policy, and a French science policy and so on down even to a Finnish science policy (in a country with half the population of Greater London, as Ms Herman points out). Europe still had (and has) the talent, but it lacked (and lacks) the structures through which to exploit that talent at anything like the same rate as America.

There have been exceptions. The European Organization for Nuclear Research (CERN), for example, and the European Space Agency (ESA), responsible for the recent Giotto-Halley encounter. CERN

began early in the 1950s as a theoretical group at the Niels Bohr Institute in Copenhagen. But even CERN took 30 years to win its first Nobel prizes, arguably because of a European disinclination to take risks. Moreover to take CERN as a model — for the European Molecular Biology Laboratory, for example — was a mistake, for high-energy physics is a very special case, in its need for enormous central machinery that simply had to be pooled to be built at all. ESA is a comparative upstart. Today it looks very healthy by comparison with the US National Aeronautics and Space Administration, but there have been many false dawns for European space industries since 1945.

It could just be that after 40 years European science is beginning to get moving. But while there is a new awareness of the need for a continental scale in science and technology, nothing will change very much until there is central European peer review and competition for funds, and total mobility of scientific labour. Until that happens, the United States and then, perhaps, Japan, will always have the edge over an old continent whose patchwork science policies Ms Herman has so ably documented. □

Robert Walgate is France correspondent for Nature.

Academic matters

Alec Merrison

Government and the Universities in Britain: Programme and Performance 1960–1980. By John Carswell. Cambridge University Press: 1986. Pp.181. £19.50, \$42.50.

THE subject of *Government and the Universities in Britain* could hardly be more important, even within the restricted canvas implied by John Carswell's subtitle. But the reader will look in vain to this slackly conceived and carelessly executed book for a serious analysis of what has happened over the past 20 years, or for guidance for the future.

Mr Carswell is, of course, unlucky in his timing, because he does not "attempt to describe events after 1978 when [his] direct experience of university-state relations came to an end"; so much of critical importance has happened since then that his book inevitably has a somewhat otherworldly air. Even so, it is an uneasy mixture of description and analysis of serious matters with personal reminiscence. I really do not think, for example, that the reader will be enlightened being told that Sir Maurice Dean, Joint Permanent Under-Secretary at the Department of Education and Science 1963–1964, was

fond of flashing his bright red braces.

So what would one look for in a book with this title? Without the subtitle — and even with it — certainly some discussion of what functions a university cannot cede to the state without losing its autonomy; indeed, discussion too of whether universities like those in Great Britain, with so much of their income coming directly from public funds, can genuinely claim to be autonomous at all. And, again, examination of the question whether universities should enjoy autonomy and self-government if they have already grabbed the Queen's shilling. Should they not be directed by democratically elected representatives and all that?

The argument for the universities' independence, as for all free institutions within the state, is the strength and flexibility this ultimately gives to the society in which they have their being. As it happens, we are, I think, particularly good in Britain at using this kind of freedom wisely, but the whole edifice depends upon convention. Universities do have their Royal Charters, but all that could be overturned in the twinkling of an eye.

In a book of this kind one would also look for an extensive and thorough account of the means of government funding other than that of the University Grants Committee (UGC), in particular that of the Research Councils. A "contrast and compare" of the block grant, blanket-style

of funding and the more specific funding of the research grant surely should have been an essential part of the book, but it seems that Mr Carswell does not think so. Since he was placed so well as an observer — first as an official at the Treasury, and at the Department of Education and Science, and, subsequently, as Secretary to the UGC — this seems a particular pity.

Then it is illuminating to view the British universities as part of an international scene. Mr Carswell makes little attempt to do so. As an example, when I was a Vice-Chancellor I learned from my American counterparts, the Presidents of the “research” universities, that the diverse funding which they enjoyed was no sure safeguard against the economic ill-wind. When such a wind blows on a national scale then there are few, if any, funding sources that are immune to it.

It would, too, have been useful to have had from someone with Mr Carswell's experience a rather more critical discussion of how well-suited the UGC is to its present tasks, particularly with Lord Croham about to report on this matter. Though, like most academics, I am a great admirer of the UGC system and those who have operated it over the years, my own view is that it is quite *unsuited* to its main responsibilities; namely, the getting and the planning of the spending of over £1,000 million a year. To start with it is grossly understaffed, both in terms of quality and quantity, to carry out such work.

The principal event of the two decades Mr Carswell is concerned with was, of course, the post-war “bulge” reaching the universities and the consequent Robbins Report and general expansion of the higher education system. Mr Carswell pays tribute to the Robbins Report as “one of the great state papers of this century”, and rightly so; one of its many virtues was that it contained so much information that if you did not like its analysis or its conclusions then you could invent your own.

Mr Carswell has a weakness for the purple — well, lavender — passage. His concluding paragraph is:

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

The state and the universities are like a discontented couple who cannot live without each other: he rich, busy, self-important, preoccupied with the office; she proud, independent and in her own opinion beautiful. The state-husband will always complain about her extravagance and inconstancy, and the university-wife will endlessly denounce his stinginess, jealousy and philistinism. Her parting words in the endlessly renewed argument will be that she knows he has a mistress — ‘and very common she is’. But they would not dream of parting: because of the children.

I think it quite wrong to interpret the relationship of the government and the universities as one of conflict. Even at the present troubled moment, as troubled as any moment in the past 30 years, I do not see the damage which is unquestionably being done to the universities by government cuts as a considered attack on the academic world. All that has happened, surely, is that we have a government committed to reducing public expenditure which has misjudged the serious long-term consequence of damage to the universities, and which believes, correctly if sadly, that the universities will not have wide public support. This last is due in part to the relatively small scale of the British university system with, in turn, so few graduates in the community; but it is due much more to the damaging antics of students in the period Mr Carswell describes, and also to the absurd misconception of university as ivory tower.

A university's first duty is indeed to preserve, create and transmit scholarship, but the modern university has little or no problem in reconciling this to its more immediate responsibilities to the community. One of my Canadian vice-cancellarial colleagues exaggerated when he said that “a university today does not know whether it is an ivory tower, a frontier post or a service station”. The fact is that it has to manage to be all three; but that, again, is not a problem discussed in this book. □

Sir Alec Merrison, Vice-Chancellor of the University of Bristol from 1969 to 1984, and a member of the Advisory Board for the Research Councils from 1979 to 1983, is a Director of Lloyds Bank plc and Chairman of the Western Provident Association Ltd, Culverhouse, Culver Street, Bristol BS1 5JE, UK.

Cutting changes in the mind

Stuart Sutherland

Great and Desperate Cures: The Rise and Decline of Psychosurgery and Other Radical Treatments for Mental Illness. By Elliot S. Valenstein. *Basic Books: 1986. Pp.338. \$19.95.*

THE mentally ill have always been treated with a savagery normally reserved for alien races. The chains that Pinel struck from his patients in the Salpêtrière were soon replaced by more insidious assaults devised by the medical profession. Within the past hundred years, mental patients have been subject to prolonged cold douches and other forms of hypothermia; they have been deliberately infected with malaria; they have been plastered with ointments to produce fevers and pustules of pus; they have had large quantities of horse serum injected into their spinal cord in order to cause meningitis; their endocrine glands have been removed; their teeth and tonsils have been taken out on the grounds that they might be harbouring infections; and female patients have had their cervixes excised. More recently, mental patients have been subjected to insulin and metrazol therapies which produce convulsions and coma, and, as a side-effect, fractured and dislocated bones. None of these treatments is known to have brought any benefit nor did they have any scientific justification, though their proponents were not slow to support them with naive and unsubstantiated theories of the nature of mental illness.

Great and Desperate Cures raises the question of how far the mutilation of the mentally ill in the name of medicine has continued into recent years. Most of the book is a history of psychosurgery, that is surgery undertaken on parts of the brain not known to be diseased or damaged with the intention of changing the patient's psychological state.

The first to undertake such surgery was Egas Moniz, a Portuguese neurologist whose ambition exceeded his talent. He invented prefrontal lobotomy, an operation in which the nerve fibres between the prefrontal lobes and the remainder of the brain are severed. It is now known that these fibres connect the prefrontal cortex to the limbic system, which is implicated in the control of emotion, but Moniz was inspired to undertake the operation by witnessing two monkeys which, it was claimed, had been made placid by lesions to their frontal lobes. Even this justification turns out to have been erroneous, since in fact one of the monkeys became more irascible after the operation. This fact was conveniently forgotten by one of

Steve Blogg

“Britain needs its universities” — but can academics expect the public to take the same view?

the experimenters, John Fulton, a prestigious American neurophysiologist, in an attempt to gain credit for initiating psychosurgery. Valenstein presents a good case for believing that Moniz's main motive was to achieve in his declining years the recognition that a rather mediocre career had otherwise failed to bring, by gambling on the success of an operation so drastic that nobody else would dare to perform it.

It was Walter Freeman, a professor at George Washington University, who in 1936 introduced prefrontal lobotomy to the United States. Over the next 30 years he was its main advocate and conducted over 2,000 operations, paying seemingly little attention to the type of mental illness from which his patients suffered. He was a buccaneer who combined coldness with considerable personal courage. He invented transorbital lobotomy, a procedure that could be carried out in a few minutes and without an anaesthetic. An ice-pick or any other instrument that came to hand was used to pierce the orbit of the eye and to mash up the adjacent parts of the frontal lobe. Freeman would perform the operation at a whim in his own office with almost no antiseptic precautions. He launched a crusade to bring this simple but unaesthetic operation to the overcrowded mental hospitals of the United States, tirelessly going from one to another to demonstrate how it was done and making deliberate use of the press to gain publicity for his claims.

Valenstein attributes the inception and growth of psychosurgery to two main causes. The first was the ambition of its proponents, ambition which in Moniz's case the Nobel Prize Committee to its shame rewarded with the 1949 prize for medicine. Second was the economic problem created by the large number of patients in mental institutions: one report even counted as an economic saving the reduction in patients caused by deaths in the operation. But a third reason must surely have been the frustration felt by doctors in the face of the sheer misery of mental illness and the knowledge that there was little or nothing they could do to alleviate it.

In fact, as soon as other methods of therapy became available, psychosurgery lost its popularity. In the 1960s it was replaced by the widespread use of electroconvulsive therapy and by the new psychotropic drugs. Moreover, the recognition that a long stay in a mental hospital hurt the patient even more than it did the tax-payer was beginning to reduce the number of long-stay patients and hence the economic argument for psychosurgery was weakened.

Valenstein avoids moral judgements, but there is one respect in which the initiators of psychosurgery were clearly guilty. Having begun, almost on impulse, to

Read the last chapter to find out how those treasured frontal lobes, supposed to be man's most precious possession, can bring him to psychosis and suicide!

PSYCHOSURGERY

Freeman and Watts

This volume inaugurates a new era in the treatment of mental disorders, a surgical era. In it the authors have assembled a wealth of observational material from eighty patients studied over the past five years. Under their skillful literary treatment these patients come to life, speak and act like human beings. Some of them you will see emerge from a distressing mental disorder even while lying upon the operating table.

Psychosurgery is new, but its foundations are laid in antiquity. One might say that from Luca trophing to modern psychoanalysis and electroencephalography, no held has failed to contribute something to this study. The authors have selected gems of theory and of observation from the past. They have also advanced some new theories concerning mental activity and the genesis of mental disorders that are, to say the least, challenging.

This is a solid factual account, illustrated with many case histories, of new... es in that excit... of the brain... the mi...

"This work reveals how personality can be cut to measure..." — blurb on the jacket of Psychosurgery by W. Freeman and J.W. Watts, published in 1942. Freeman himself wrote the copy.

use an operation that irreversibly damages the brain, they continued to use it on an increasingly wide scale without making any systematic evaluation of its effects. It became apparent that lobotomy could cause massive changes in personality: these changes were unpredictable in that it could have a calming effect but could sometimes produce extreme irritability. Many patients were rendered highly distractible, and also less flexible in carrying out a task. Many became apathetic and lacking in creativity or emotional spontaneity.

Despite these difficulties, none of the pioneers of the operation and few of their successors conducted controlled outcome studies. They failed to record adequately the pre-operative histories of the patients; they did not use a control group which had not undergone the operation, although such a group is particularly important in assaying the therapeutic effects of any treatment for mental illness because spontaneous recoveries are frequent; they appointed themselves both executioner and judge, since the post-operative assessments were conducted almost entirely by the doctor who had been responsible for the operation; and they failed to break their cases down by type and severity of illness. In these respects, however, the psychiatrists were little worse than their colleagues in other specialities, for outcome research in medicine as a whole has been (and often remains) a disgrace. Random control trials were introduced in agriculture and psychology at the turn of the century, but the first study of this sort in medicine did not take place until 1953. Only a few years ago an attempt to set up a careful study of lobotomy in Britain, which had the support of the Medical Research Council, foundered because the psychosurgeons refused to co-operate. One is reminded of the psychiatrist who, on being presented with a design to test outcome, pointed at one of the groups and asked what it was for. On being told that it

was a control group, he grunted and said, "We don't need that, take it out. I'm in control here".

Regrettably Valenstein does not present a systematic review of the effects of psychosurgery. Some reports, particularly those of Freeman, suggest that it may help some people and two independent studies have reached the same conclusion. But its effects appear to vary unpredictably from person to person and there is no question that it can do terrible harm. Unfortunately it was not always used as a last resort and there are pathetic cases in which it has turned previously cheerful adolescents into zombies.

The question of the effects of psychosurgery is by no means an academic one, for several hundred such operations a year are still being performed in the West. Those on mental patients rarely receive any publicity, but in the United States and elsewhere prisoners have received psychosurgery in order to reduce violent behaviour or rid them of sexual compulsions, and several operations have been undertaken on recalcitrant children. Some of these operations were performed with consent granted in return for parole; others, notably those on children, without consent.

Valenstein's concluding chapter is too short and too hurried. He presents his story as a cautionary tale and points to the rapid increase in the United States of coronary by-pass operations, many of which are being carried out on patients who, according to a recent large-scale study, can derive no benefit from them. Careful research is clearly not a complete answer, because it can always be ignored by over-enthusiastic or avaricious surgeons. Valenstein himself proposes that new surgical procedures should be subject to the same restrictions as new drugs: despite the outcry to be expected from surgeons, this would appear to be only common sense. Such procedures would only be carried out on a large scale after research (initially on animals) had demonstrated that their benefits outweighed their risks. Valenstein is aware that surgeons would argue that such restrictions would deprive some people of an operation from which they might benefit, but they would also prevent people who would not benefit from being mutilated; moreover, the opinion of a committee is likely to be more dispassionate than that of the surgeon who wishes to practise the operation.

Despite Valenstein's puzzling failure to deal fully with the effects of psychosurgery and the ethical issues it raises, his well, often very well, written account of its inception and hey-day is both fascinating and salutary. □

Stuart Sutherland is Director of the Centre for Research on Perception and Cognition, University of Sussex, Brighton BN1 9QG, UK.

Life of learning with the fortunate few

A.P. French

Bird of Passage: Recollections of a Physicist. By Rudolf Peierls.
Princeton University Press: 1986. Pp. 350. \$29.50, £21.20.

ANY physicist who did not participate in the flowering of quantum physics during the 1920s and 1930s (and today of course that means all but a handful) must feel some sense of envy, mixed with a touch of awe, towards those who did. It was, truly, a heroic time, during which the secrets of the atom and then the nucleus surrendered to a band of brilliant theorists and experimenters. And, largely because the community of physicists was then relatively small, it took place in an atmosphere of scientific internationalism that has probably never been surpassed. Rudolf Peierls was one of those fortunate few, and in his autobiography he writes of his experience of it, during which his own distinguished career was launched.

Peierls, born in Berlin in 1907, entered the university there in 1925 and attended the theoretical lectures given by Max Planck, who started it all in 1900 with his discovery of the quantization of energy. It should have been an exciting beginning, but, to quote the author, "Planck's lectures were, I think, the worst I have ever attended". That first year did, however, have its compensations. Peierls learned about Bohr orbits and the quantum of action in a lecture course on X-ray physics, given by Walther Bothe. He also attended lectures by Walther Nernst, whom Peierls describes (and I cannot resist quoting him) as "a great physicist of small stature and an even smaller sense of humility" — but nonetheless a very good lecturer.

Better things were in store. In the next year he moved to Munich, where Arnold Sommerfeld, extender of Niels Bohr's atomic model and supreme teacher of the theoretical physics, was the revered "Geheimrat". Here, Peierls's eyes were opened to the newest of the new. Werner Heisenberg and Erwin Schrödinger, in their different (but, as it was later shown, equivalent) ways, had just published completely fresh approaches to the problems of atomic dynamics; and the formidable Wolfgang Pauli, a former student of Sommerfeld, had introduced his famous "exclusion principle", according to which no two electrons could occupy the same quantum state. Physics was undergoing its greatest transformation since Faraday and Maxwell had unified electricity and magnetism. In 1928, in pursuit of the new knowledge, Peierls went to study under Heisenberg in Leipzig, and then, in 1929, to join Pauli in Zurich. At the end of that year he obtained his doctorate and became Pauli's assistant. But his horizons

were still expanding. In 1930 he visited Bohr in Copenhagen and then, in an exciting and benignly fateful trip, attended a conference in Odessa, where he met Eugenia Nikolaevna Kannegiser from Leningrad. Little more than six months later he returned to Leningrad to marry her.

With the rise of Hitler in the 1930s, physics and politics became painfully intertwined. First came the exodus of refugees



Early days — Rudolf Peierls at an informal meeting in Leipzig, 1931.

from Germany and Austria. At this time, Peierls and his wife happened to be in Britain, thanks to a Rockefeller fellowship, but it meant that there was no going back. The ultimate result was that, from 1937 to 1963, Rudolf Peierls was professor, and leader of theoretical physics, at the University of Birmingham. During this quarter-century he built up an outstanding research school, notable alike for the catholicity of its physics and the diversity in national origins of its students.

The second major development — closely related to the first — was of course the atomic bomb. It is amazing to contemplate how many of the leaders in that enterprise had been driven out by the Axis powers. Amongst the physicists, one can immediately think of Hans Bethe, Enrico Fermi, Otto Frisch, Victor Weisskopf and, of course, Peierls himself. Otto Frisch, who in Niels Bohr's Institute in 1939 had been the first to observe directly the disruption of uranium nuclei through fission (and who had given the process its name) joined Peierls at Birmingham a few months later. Early in 1940 they wrote a famous joint memorandum, spelling out in some detail the possibility of making a "super-bomb" with only a few kilograms of separated uranium-235. The idea of making an explosive based on fission had been in the air since fission was first discovered, but given the low "abundance" (only 0.7 per cent) of the easily fissionable

^{235}U in natural uranium, the belief in this possibility was about to expire when the Frisch-Peierls memorandum was sent to Sir Henry Tizard, then chairman of the Air Defence Committee. The result was the establishment, under G.P. Thomson, of the curiously named "MAUD Committee" (which, as Sir Rudolf has since remarked, was intended to be meaningless, but could have been interpreted as "Military Applications of Uranium Disintegration" — altogether too revealing!).

Two central chapters on the wartime years — divided, for Peierls, between Britain and the United States — give an excellent picture of the scientific and human context in which the atomic bomb was made. Whatever its darker implications, this was for the participants an exciting time; particularly so, perhaps, for the small group of British scientists who, like Peierls himself, joined the American project at Los Alamos, where the bomb was completed and assembled. Looking back, the youthfulness of the leading figures in this enterprise is little short of astounding. Oppenheimer was appointed Director of Los Alamos in 1942, at the age of 38, and Hans Bethe became head of the Theoretical Division in 1943, at age 37. Peierls himself was also 37 when he went there, succeeding Chadwick as head of the British mission. Also at Los Alamos, of course, was Klaus Fuchs, who had been brought into the project by Peierls himself in 1941 and, as Peierls describes, must have begun his leaking of atomic secrets to the Soviet Union almost immediately.

After the war it was Birmingham again, for nearly 20 years. It was a time of steady development, as new ideas in many different areas of physics made their appearance. Peierls, the generalist, welcomed them all, in what seems to have been a happy association of staff and students. It is noteworthy that a whole chapter in his book is entitled "Teaching"; he took his responsibilities seriously. After Birmingham it was Oxford — a decade of developing a strong department of theoretical physics. Then, in 1974, came a well-earned retirement.

I should have preferred to have seen somewhat more of Sir Rudolf's views on the state of physics, and less of personal anecdote. But clearly he has written the book he wanted to write, and in the process has made a enjoyable addition to the living history of twentieth-century physics. Moreover, he projects an attractive image of a talented but genuinely modest person who has made a notable contribution to the advance of education and research in physics, especially in Britain.

A.P. French is a Professor in the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. He was a junior member of the British "tube alloys" project, and of the British group at Los Alamos during the war.

Irreversibility and the resurrection of the dead

Ilya Prigogine and Yves Elskens

Une Incertaine Réalité: Le Monde Quantique, la Connaissance et la Durée. * By Bernard d'Espagnat. *Gauthier-Villars, 17 rue Dénys-Dumoncel, BP 50, 75661 Paris Cedex 14, France: 1985. Pp.310. Pbk FF98.*

BERNARD d'Espagnat is well-known as the author of several excellent monographs on contemporary physics; his *Conceptual Foundations of Quantum Mechanics* (Addison-Wesley, 1976) and *Nonseparability and the Tentative Descriptions of Reality* (*Phys. Rep.* **110**, 202–264; 1984) have become classics in their fields. Readers will therefore open his new book with anticipation, not least because its subtitle, *Le Monde Quantique, la Connaissance et la Durée*, promises an original discussion of such fundamental contemporary questions as causality and irreversibility.

d'Espagnat defends here the thesis that today's physics requires two irreducible notions of reality: on the one hand, an "independent" reality, and on the other an empirical reality. This is an idea which was also to be found in his preceding book, *A la Recherche du Réel* (Gauthier-Villars, 2nd Edn 1981). Using arguments originating in both epistemology and the natural sciences, d'Espagnat discusses this concept in detail but in terms that are accessible to the general reader. As would be expected, the "independent" reality is timeless while irreversibility appears only in empirical reality.

The thesis of two opposing realities is not new, nor is the response of most physicists to it. One is reminded of Léon Rosenfeld's "lay sermon", published in his *Selected Papers* (Reidel, 1976) in which he likened the attitude of many theoretical physicists to irreversibility to that of the Athenians towards Paul's preaching:

they quietly let him talk — until he came to the resurrection of the dead. Then they laughed at him and went away, saying: 'We'll listen to this another time'.

Similarly, physicists tend to find better things to do once irreversibility is mentioned: they know that irreversibility belongs only to the phenomenological world and (Rosenfeld again):

they assume strict reversibility in 'another world' — the world of atoms, of which until recently they had no better knowledge than the

apostle could have had of paradise or the other place.

Be this as it may, the dualistic conception of reality does put physicists in a difficult position. Irreversible processes are seen to be of growing importance, not only on the phenomenological level, but also on that of fundamental descriptions (examples are the quantum theory of measurement, unstable elementary particles and cosmology in the early stage of the Universe). This trouble is aggravated when one notices that quantum processes cannot be described in terms of macroscopic space-time concepts when repeated measurements are involved (see, for example, B. Misra and E.C.G. Sudarshan's discussion of the quantum Zeno paradox which appeared in *J. Math. Phys.* **18**, 756–763; 1977).

In the book, d'Espagnat makes frequent (and penetrating) reference to the work of our group in Brussels, but in his view this work establishes irreversibility only as an "instrumentalistic", phenomenological notion of reality. It is true that our understanding of the microscopic meaning of irreversibility, while dating back to Boltzmann, is still in an exploratory stage. However, in recent years a rigorous link has been developed between highly unstable classical dynamics (for example in Kolmogorov systems) and irreversibility. It turns out that the classical pictures of point dynamics, and of quantum theory, undergo basic modifications when the processes of observation and measurement are taken into account for unstable dynamical systems. This instability precludes the usual description in terms of trajectories and opens the way to a non-local, statistical description including irreversibility. The finite precision of observational data is not, in our opinion, specific to human measurement and must be included in every consistent theoretical scheme. This development tends to support the duality of reality suggested by d'Espagnat; curiously, he himself, at the end of the book, seems to lean towards acceptance of the notion of some deeper "réel vrai".

It is fascinating to see how today's physics echoes the great themes of Western philosophy; after all, d'Espagnat's two realities are close to the two "worlds", noumenal (intellectual) and phenomenal, introduced by Kant. More recently, of course, C.P. Snow examined the "two cultures". Anyone interested in what unites them at the deepest level will find d'Espagnat's book to be well worth reading. □

Ilya Prigogine and Yves Elskens are in the Service de Chimie Physique II, Université Libre de Bruxelles, Campus Plaine CP 231, Boulevard du Triomphe, B 1050 Brussels, Belgium. Ilya Prigogine is also Ashbel Smith Regenia Professor in the Department of Physics, University of Texas at Austin, Austin, Texas 78701, USA.

*An Uncertain Reality: The Quantum World, Knowledge and Time.

Physics in concert

Pedro Waloschek

QED: The Strange Theory of Light and Matter. By Richard P. Feynman. *Princeton University Press: 1986. Pp.158. \$18.50, £13.40.*

THE message of this book is clear: Richard Feynman considers that the principles of quantum theory, and in particular of quantum electrodynamics (QED), should now be revealed to the world at large. Here he explains QED simply and without mathematics. Feynman, who earned the Nobel prize in 1965 for his contributions to QED, shows that this theory is as natural as numbers and that it describes 99 per cent of what happens with incredible accuracy.

In the past we used numbers to describe the world about us — and thereby made possible the myriad activities based on them — but now we use arrows. There is no magic about them and Feynman shows how easy they are to work with. By the end of the first chapter the reader is brought to understand that even the simplest phenomena, such as the reflection of light in a shop window, can only be described correctly using these little symbols. Newton's particle theory of light, or Maxwell's wave-theory (or both combined in some funny way), will never provide a correct prediction of, for example, the results of an experiment on reflection — if it is carried out with sufficient accuracy. The arrows, which move like the hands of a clock as time passes, are simple tools. They do not tell us why things happen, but how to describe them.

The book is made up of four chapters, which correspond to each of Feynman's Alix G. Mautner Memorial Lectures given at the University of California, and which are like four parts of a concert. First comes an *adagio*, then an *andante*. Next is the *crescendo*, followed by the overwhelming *finale*, which probably only those with prior knowledge will ever understand. But this is part of the game. First Feynman tells us how simple it all is. In Chapter 2 dawns the realization that more details must be taken into account. In Chapter 3 things become more complicated still, but the attentive reader remains convinced that everything, including the minutiae of atomic structure, can be calculated by using the arrow principle. Finally, Feynman tries to summarize our entire knowledge of particle physics in exactly 28 small-format pages — a valiant, hopeless enterprise, but resounding all the same.

The arrows are the probability-amplitudes which have been used by physicists over the past few decades for all their computations. They correspond in mathe-

mathematical language to complex numbers. But most physicists have never tried to find an intuitive way of thinking about them; we were always told that it could not be done. Feynman demonstrates that it can, and goes on to introduce another widely used tool in physics, his own Feynman graphs. In a simple and intuitive manner he shows how these diagrams can be "imagined", without understanding what is behind them but getting a good feeling for how they work. Years ago I heard a seminar by Feynman, in which he observed that many physicists used the graphs in much the same way that our ancestors would consult animal entrails for portents of the future. Now it seems that a new understanding has taken over; the graphs are taken as seriously as the arrows, and together they frame the world of present-day physics — the only valid description of nature we have.

Feynman's lectures must have been marvellous and they have been turned into an equally entrancing book, a vivid introduction to QED which is leavened and enlivened by his wit. Anyone with a curiosity about physics today should buy it, not only to get to grips with the deepest meaning of quantum theory but to possess a slice of history. □

Pedro Waloschek is a senior scientist at DESY (Deutsches Elektronen Synchrotron), Notkestrasse 85, D-2000 Hamburg 52, FRG.

Schwinging through relativity

Abraham Pais

Einstein's Legacy: The Unity of Space and Time. By Julian Schwinger. W.H. Freeman: 1986. Pp.250. £15.95.*

THE scientific community appreciates quite well the importance of conveying to a wider audience what science is about. So high are the demands of doing research, however, that few of those active in the field are willing to respond to this need. Doing science and writing about it are actually two distinct activities requiring distinct skills. Nevertheless I like to think that in principle the scientific craftsman is best qualified for popularization. All the many subtleties of any specific scientific subject cannot, in fact should not, be presented to those eager to obtain a first acquaintance with it. Awareness of these subtleties by the popular science writer will, however, temper his style, preventing him from couching his account in those overheated terms and exalted phrases so often used by those who do not really know what is going on.

Julian Schwinger is a pro, particularly well known for his seminal contributions to quantum electrodynamics, for which, with Feynman and Tomonaga, he was awarded the 1965 Nobel Prize in physics. Those as his other writings are marked by consummate mathematical skill. I must confess to considerable astonishment upon receiving his popular account *Einstein's Legacy*. That, I said to myself, is not the Schwinger I know well. I am happy to report, however, that his book makes for delightful and instructive reading. Technical demands on the reader do not exceed the most elementary algebra.

In the book, parts of Einstein's rich and varied legacy, on quantum theory, on Brownian motion, are touched on only in passing. As the subtitle *The Unity of Space and Time* indicates, the main thrust is here on relativity theory. The opening chapter "A Conflict Brought to Light" (several of the chapter titles are puns) serves as a prelude to the subject. It describes Michelson and Morley's failure to detect the speed with which the Earth drifts through the ether, that ubiquitous medium, for so long erroneously supposed to be necessary for explaining how light waves travel through space. It is explained why this negative result is of such positive value. Here and elsewhere the author acquaints the reader with the issues by first considering a more familiar analogy, in this case the problem of find-

ing the speed of aircraft moving along with or perpendicular to the jet stream.

The next chapter, "Marking Time", introduces a variety of clocks: pendulums, vibrating atoms, beams circulating in cyclotrons. Soon we are in the midst of the results of relativistic measurements, so different from everyday expectations because of the relativity postulate that the speed of light is unaffected by the motion of the observer. Time intervals dilate, lengths contract — not really, but as seen by an observer at rest who sees, respectively, moving clocks and moving rods. Here, as elsewhere in the book, subtle concepts are elucidated by simple



Showbiz meets physics — Einstein and Charles Chaplin in Hollywood, 1931.

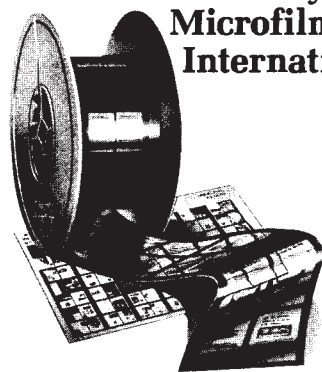
pictures. Thus the notion of causality is explained by describing the sequence of events at a race track. Cartesian coordinates are introduced by taking the reader on a walk through midtown Manhattan.

A chapter on $E = mc^2$ deals with the meaning of that relation in terms of conservation laws and with its applications to (in order of practical importance) chemical, atomic and nuclear reactions, including fission and fusion.

All this, so far, is special relativity which relates experiences of observers in uniform relative motion. As Einstein once said, that theory is child's play compared to general relativity which deals with non-uniform motion. It is particularly gratifying that the subject's intricacies, including glimpses of non-Euclidean and Riemannian geometries, and the theory's early successes, the bending and red shift of light, and the precession of the planet Mercury's perihelion are so well conveyed here in simple language. Altogether, this well printed and pleasingly illustrated book is an ideal gift for the curious non-expert. □

Abraham Pais is a Professor in the Department of Physics, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

nature is
available in microform
from University
Microfilms
International.



□ Please send information about these titles:

Name _____
Company/Institution _____
Address _____
City _____
State _____ Zip _____
Phone () _____

Call toll-free 800-521-3044. In Michigan, Alaska and Hawaii call collect 313-761-4700. Or mail inquiry to: University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106.

*Published in the United States as a volume in the Scientific American Library.

What's wrong with being neutral?

Theodore K. Rabb

Climate Impact Assessment. SCOPE Report 27. Edited by Robert W. Kates, Jesse H. Ausubel and Mimi Berberian. Wiley:1985. Pp.625. £59, \$100.

SCIENTISTS are understandably shy of making doomsday predictions. Even when their investigations uncover what some might consider frightening possibilities, they incline to caution. The calculations are uncertain, extrapolations are full of pitfalls and nobody is sure what the consequences for humanity of specific events might be. No area of science is more fraught with such hesitations than the study of the effects of climatic change.

The editors of *Climate Impact Assessment* have put together a breathtaking survey of what has been accomplished in over three decades of research in four main areas: overall models and frameworks for assessing climatic impacts; the impact in specific sectors such as agriculture and water resources; the impact on particular societies and their response; and the construction (often by modelling) of integrated assessments. But the emphasis throughout is on neutral interpretation. For example, because short-term changes are much larger and therefore mask long-term changes, it is said on p. 42 that one should withhold judgement on the rise in the concentration of carbon dioxide in the atmosphere — this, when 1.5×10^{10} tons are being added each year, and an almost equivalent tonnage of ozone is being lost (p.505). Similarly, neither observed changes nor their variability are "significant" (p. 40) — to whom? And narrative descriptions of climatic disasters, containing such rhetoric as "crippling", are to be avoided, because "the unwary reader" may not discern the absence "of a precise framework for gauging the importance of climatic impact" (p. 543). The measured tone, the rap on the knuckles for the few unruly children who do make urgent noises — this is the seventeenth-century heritage of science as an escape from emotion taken (at least for this field) too far.

References to the level of carbon dioxide in the atmosphere recur throughout the 22 essays in this book. But all of the authors treat it as a valuable statistic, as a stimulus to research or as a fascinating academic question. Thus (p. 509):

An increase in the average temperature by 3 or 4°C could lead to the beginning of an irreversible melting of glaciers. What will the properties of the new state of equilibrium of the biosphere be like; will they permit the existence of man? We do not know.

Those who have suggested they do know, and have foreseen terrifying dislocations as the Earth's landmass shrinks dramatically, tend to be relegated to the fringes of scholarship. They are scarcely mentioned in this book, as if they might somehow tarnish the respectability of the field. One has the feeling that perhaps everyone is hoping that the dire predictions will balance out — that the greenhouse effect might be counteracted by the cyclical return of an ice age — so that what Thomas Kuhn called "normal" science can proceed undisturbed.

Yet the basic facts are inescapable. Given the comment in the quotation above about a 3 or 4° C temperature rise, what is one to make of the 2° C rise that has taken place in Indiana over the past century (p. 61)? One answer emerges unmistakably: recognition of a potential problem is crucial if a society is to have any interest in addressing it. A 30 per cent increase in summer rainfall recently went unnoticed by the citizens of St Louis (pp. 324–326). People do make small adjustments, of course; but what is likely to follow if — as has happened — a fundamental but imperceptible shift takes place, such as a displacement of the northern limit of a particular crop's viability more than 300 km southward over the course of a century (p. 365)? Do we throw more chemicals at it? Relocate or retrain the populations of entire regions?

None of this is to suggest that there is no real concern about the issues in the scientific community. The Carbon Dioxide Information Center at Oak Ridge National Laboratory is one of a number of organizations that keeps trying to draw our attention to the need for both research and action. But it is the former that takes precedence over the latter, and by a long way. A briefer version of the thorough review of the scholarly literature in *Climate Impact Assessment* appeared in

Kellogg and Schwere's study of the consequences of the build-up of carbon dioxide (*Climate Change and Society*, published by the National Center for Atmospheric Research in Boulder in 1981). They reached more dramatic conclusions, and linked their call for more research with suggestions of strategies for dealing with the problem. But the practical results have been nil — a response that recalls the Sherlockian dog that failed to bark.

Lucid, precise and abundantly informative though these essays are — I noticed only one misprint that might mislead: 200 on p. 107 should be 2,000 — they do not take the process of public education beyond Kellogg and Schwere's book. Perhaps that requires a different kind of effort which, though doubtless likely to attract frowns, is in fact no less admirable than the meticulous overview of research issues and conclusions that these authors provide. There may be uncertainties aplenty, but the expectation of a major warming during the twenty-first century now seems to be almost universally shared. In addition to refining the tools by which this process should be measured and assessed, scholars now have the responsibility to specify more urgently the dislocations that are likely to ensue. This book is a foundation for that step: it identifies the topics and findings on which all future work must build. But one has to regret that the learned and superbly qualified authors who produced it have held back from attempting to offer the sharp and unambiguous conclusions that could have served to educate a wider public. □

Theodore K. Rabb is a Professor in the Department of History, Princeton University, Princeton, New Jersey 08544, USA. He is co-editor of *Climate and History* (Princeton University Press, 1981).

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Within Saturn's rings — detail of a painting by William K. Hartmann (the original is in colour). The illustration is reproduced from Out of the Cradle: Exploring the Frontiers Beyond Earth, by Hartmann, Ron Miller and Pamela Lee, published by Workman, New York. Price is hbk \$19.95; pbk \$11.95.

Picture the planets

Michael H. Carr

Planetary Landscapes. By R. Greeley. Allen & Unwin: 1985. Pp. 257. \$44.95, £24.95.

DURING the past 20 years spacecraft have been sent far and wide on exploratory missions throughout the Solar System; among the planets, only Neptune and Pluto have so far escaped close inspection. The result has been an explosive growth in our knowledge of what the various bodies of the Solar System are like, and how they evolved to their present state. This burst of scientific activity is now being reflected in the appearance of a stream of books. The knowledge is so new, the subject so large and the potential readership so broad, however, that there is as yet little overlap between the successive publications.

Planetary Landscapes concentrates almost exclusively on the surface morphology of the solid planets and moons, dealing primarily with Mercury, Venus, Mars, the Earth's moon, and the moons of Jupiter and Saturn. The book describes the physiography of these bodies and discusses briefly the clues that their surface morphology provide about the nature and timing of the various processes that have operated on each of them to produce the different configurations we see.

Ronald Greeley, who has participated in the planning and execution of several planetary missions, and who has done research on the effects of impact, volcanism and wind on different planetary surfaces, is well qualified to write a book such as this. In his two introductory chapters, he describes the history of our exploration of the Solar System and some of the techniques used in planetary geology, particularly digital image processing. As in most chapters, the story is told mainly through illustrations and tables. There then follows a chapter which discusses the four principal types of process that affect the morphology of a planet's surface: impact, tectonics, volcanism and gradation. The rest of the book is a tour of the various planets and moons, most chapters being organized around a descriptive section and a discussion of the extent to which each of the four types of process have operated on the body being considered. The approach is qualitative, narrowly focused on the information embedded in the landscape, and the text is only lightly referenced; as the author says in the preface, the references are not meant to be exhaustive but rather to provide a springboard to further reading. In this regard the tables of special journal issues and NASA publications on the planets are particularly useful.

The main strength of the book is its illustrations, over 250 of them. Some 80% of these are spacecraft images, the rest being maps, diagrams and pictures of geological features of the Earth that have some particular planetary significance. I know of no other publication that provides such a rich collection of planetary photographs; there is hardly a page without an illustration, and many cover almost the entire 11" × 8½" format. The figures are well captioned and well annotated, the spacecraft images are identified by number and information is provided on how copies might be obtained.

If I do have a quibble, it is that the

quality of reproduction is somewhat disappointing, but this is a minor criticism given the many spectacular views that are presented. The text, however, is rather basic, and seems to be intended neither for the professional nor for the casual reader; rather the book appears to be designed as an illustrated account for the serious layperson or as an introductory text on planetary geology. I anticipate that it will be used widely in both of these roles. □

Michael H. Carr is at the US Geological Survey, MS 946, 345 Middlefield Road, Menlo Park, California 94025, USA. He is author of *The Surface of Mars* (Yale University Press, 1982).

Mapping decline

Peter J. Smith

Down to Earth: One Hundred and Fifty Years of the British Geological Survey. By H.E. Wilson. Scottish Academic Press: 1985. Pp. 189. Pbk £9.50.

UNSUSPECTING readers who, seeing no further than the subtitle, imagine this to be an official, rigorous or even entirely serious history of the British Geological Survey's first century and a half will be sadly disappointed. Suspecting members of the geological community who have heard rumours to the effect that the book is one long string of saucily amusing anecdotes will be no less disappointed. As far as the (really rather few) scurrilous stories are concerned, Wilson comes over for the most part as the Alexei Sayle of geology, a purveyor of comedy with the funny bits left out. Much more interesting is his self-ennoblement as Lord Gnome of Keyworth. Tales of BGS geologists relieving themselves against trees are far less important than the inside story of how government and the Natural Environment Research Council have relieved the Survey of its vitality.

Each of Wilson's 18 chapters takes an insider's look at a particular aspect of the Survey's work, generally from a historical perspective but with comparatively short shrift being given to the first 110 years. Indeed, it soon becomes clear that Wilson's chief concern is the spectacular rise and fall of the BGS during the postwar period. Before 1939 the Survey was almost entirely a geological mapping agency, albeit one often in the charge of megalomaniac Directors more interested in self-advancement than the industrial needs of the nation. As government interest in science grew, however, the organization changed in both character and size as it absorbed new techniques, adopted new goals, and appointed new people unhindered by a century of tradition.

Wilson has done one of his finest jobs in

delineating the myriad organizational changes that this postwar expansion apparently necessitated. The extent of the formation, growth, transmutation, decline and demise of groups, units, departments, divisions, bureaux, branches and teams over the past 20 years has been truly bewildering. It would be comforting to see it all as the mark of a vigorous organization perpetually regrouping to meet the demands of an advancing industrial civilization; unfortunately, whilst that may be what it seemed like at the time, it is all too easy in retrospect to view the furious reorganization as a communal fit induced by a ruling establishment chronically unable to decide what to do for the best.

In fact, as Wilson so ably explains, the government decided what to do for the worst, imposed the Rothschild principle in 1972, and thus set the BGS on the road to ruin. By the late 1970s the Survey had become so far removed from its basic function that it was relying on vulnerable short-term contracts for 80 per cent of its budget and was a constituent part of a research council that, dominated by members of a different discipline, all too clearly had little expertise, and even less interest, in the Earth sciences. When the public spending cuts came, the recipe for disaster was complete. The contractor departments of government reneged on many of their commitments and brought the BGS to crisis.

I cannot help feeling that the crisis might have been rather less acute if only the BGS had done much more in the past to gain friends by explaining what it has been about. *Down to Earth* can be regarded as a belated attempt to do just that, for in the course of explaining the Survey's problems it also provides a wealth of information on past and present activity, both in Britain and elsewhere. Whether its irreverent tone will endear it to those who are supposed to be looking after the organization's interests is quite another matter. □

Peter J. Smith is Reader in Earth Sciences at the Open University, Walton Hall, Milton Keynes MK7 6AA, UK.

How to get heard above the noise

David R. Rosseinsky

The Chemist's English. By Robert Schoenfeld. VCH, Weinheim, FRG:1985. Pp.173. DM 42, \$16.95.

ENGLISH is nowadays *the* lingua franca — anybody not perceiving all the lovely nuances of this cliché should read no further. Dr Schoenfeld's little book on the use of the language has come to the right reviewer, for how many people have had a research paper sent to four referees because of disagreements on the prose alone? (Comments ranged from "highly elegant exposition" to "turgid".) I regret having discarded substantial piles of such anonymous invective, which, as a collection, would have made for entertaining reading; after grudging concession of a major scientific point in a recent, treasured tirade, it was stated that "The gratuitous French phrase on p.6 [*faute de mieux*] should be removed — its present presence in a paper with such a poor standard of English is sheer exhibitionism" (*sic*). Failing to find a Latin equivalent, and being ever compliant, I substituted *pis aller* (not, I fear, exactly the same thing). Dr Schoenfeld does warn against flaunting foreign phrases, which — *vide supra* — can become cheap avenging targets for referees whose fox one has unwittingly shot.

The author is editor of the *Australian Journal of Chemistry*, and has collected articles of guidance aimed principally at contributors and potential contributors. The book might well serve as a prescribed text for PhD students. Dr Schoenfeld admits that he cannot say what is "good" writing, and that he only edits to achieve an "acceptable" paper free of mistakes, but he proposes that a good paper is one written to yield its information to the reader in the shortest possible time. This dodges the question of the level of the reader's expertise, and clearly the writer must judge. In my opinion a paper on dihalobenzenes and potassium amide in ammonia, written in verse by Bunnett and Kearley (*J. Org. Chem.* **36**, 184; 1971), serves as a model of clarity and readability simply because of the necessarily careful choice of words and phrasing. The failure of chemical verse to catch on is doubtless in part a result of the difficulty of such composition.

Early on, Dr Schoenfeld distinguishes Chemist's English (CE) from Standard Written English (SWE) which is nice and helpful but introduces just the unnecessary sort of neo-acronym specifically warned against later in the book. As cited, Chemist's English is a degraded tongue

whenever its usage departs substantially from Standard English; in my view the only allowable difference is a more impersonal style, and even then, only within limits. American usage is noted as allowing "we" and sometimes "I", rare in British scientific literature, but obviating the clumsy passive. Special or specific meanings are a part of scientific writing, but (to digress) did the once-proposed barbarism of "specific" = "divided by mass" ever survive? It was meant to complement "density" meaning "divided by volume", but this latter usage, having merit, has indeed happily coexisted with its English progenitor.

The book presents views from Gowers, Fowler, R.S. Cahn, eminent antipodean scientists and one T.M. Bernstein (of whom it is said, perhaps wistfully, "His writing is brisk, lucid and completely free of headmasterly condescension"). Unrelated "it"s, unrelated participles, and clumsy constructions, punctuations and styles, are all pinned down and analysed. Degraded English is opposed but not always vigorously: to give up with mere handwringing on the idiot use of "facile", usually together with "synthesis", seems to me to lack moral fibre. What is wrong with "convenient", which does not call for the instant rejection Dr Schoenfeld confers on the other alternatives?

A sometimes contrived jokiness detracts from the book's message, as does the use of colloquialisms such as "glitch", from computerbabble, whereas the best teacher would have been strict example. Further weakening of didactic authority follows from persistent use of "anthropomorphization" to which almost any circumlocution would have been preferable. And I absolutely chickened out (why should I abjure colloquialism?) of reading the chapter entitled "A Chemical Analysis of the English Sentence" — the author believes that the average reader's knowledge of chemistry exceeds that of their English, especially if the reader is under 30 and educated in an English-speaking country. Well, well! In another respect Dr Schoenfeld gets gamma-minus for an ignorant onslaught on the presentation of units in the quantity calculus. Most of these criticisms, of course, are statements of personal preference.

In my favourite article on aspects of the scientific literature, published in the *Times Higher Educational Supplement* on 8 February, 1980, Graham Hills wrote that "The signal-to-noise ratio declines yearly not because the signal is becoming weaker but because the noise is becoming deafening". The concerns expressed by Dr Schoenfeld should help moderate the din. □

David R. Rosseinsky is a Reader in the Department of Chemistry, University of Exeter, Exeter EX4 4QD, UK.

NEW RELEASES — —FROM PLENUM

DIOXIN, AGENT ORANGE

The Facts

by Michael Gough

"Gough has provided the best account yet available on the implications of dioxin. . . . His study will be in the forefront of dioxin literature, past, present, and future."

—Kenneth R. Feinberg,

Special Master, Agent Orange Trial

0-306-42247-6/290 pp./ill./1986

\$17.95 (\$21.54 outside US & Canada)

THE JOY OF SCIENCE

Excellence and Its Rewards

by Carl J. Sindermann

"... a compendium of tips that may prove useful to some about to embark on a career of scientific research in academic, government or industrial laboratories. The advice is, in general, sound and presented in an organised manner. . . ."

—*New Scientist*

0-306-42035-X/278 pp./ill./1985

\$16.95 (\$20.34 outside US & Canada)

THE DA CAPO SERIES IN SCIENCE

Outstanding Biographies of Scientific Pioneers

An exciting new paperback series featuring reprints of works that have proven to be historically significant and scientifically sound. The first two volumes:

THOMAS A. EDISON

A Streak of Luck

by Robert Conot

0-306-80261-9/608 pp./35 photos
\$13.95

LOUIS PASTEUR

Free Lance of Science

by René Dubos

0-306-80262-7/462 pp./25 photos
\$11.95



Plenum Publishing Corporation
233 Spring Street
New York, N.Y. 10013

In the United Kingdom:
88/90 Middlesex Street
London E1 7EZ, England

Chemists' conflict

Alastair Hay

The Poisonous Cloud: Chemical Warfare in the First World War. By L.F. Haber. Clarendon: 1986. Pp.401. £35, \$59.

Chemical Warfare. By Edward M. Spiers. Macmillan, London/University of Illinois Press: 1986. Pp.277. £27.50, \$24.95.

"WHAT did you do in the war?" is not a question any son of Fritz Haber would have needed to ask. Nor, indeed, would the answer necessarily be what a child of his would have liked to hear. For Haber had another, more dubious paternal role. He is the acknowledged father of chemical warfare, the pioneer of the use of chlorine in Flanders in 1915.

Haber never apologized for his part in chemical warfare: he was a patriotic German doing his best for the Fatherland, and his only regret seems to have been that this method of warfare wasn't more successful. According to his son Ludwig, the author of *The Poisonous Cloud*, Haber's opinions on chemical warfare went "unreported" and his two lectures on the subject are "disappointing"; but clearly he felt his brainchild had failed "in other hands" — those of the poor bloody infantry.

As befits a superb scientist (he was awarded the 1919 Nobel prize for chemistry, for his role in the Haber-Bosch process for making ammonia), Haber's suggestion to use gas was the result of a perceptive analysis of trench warfare in the First World War. Something was needed to break the stalemate. Gas, released from cylinders — but not projectiles which were forbidden under the 1899 Hague Convention — would do it. The German High Command, however, was not convinced. In consequence, on 22 April 1915 when gas was first used in anger, they failed to provide sufficient troops to enable a decisive break to be made in the Allied lines. Haber's view of the event is that it was a sort of experiment, poorly conducted. Not enough gas was released, and the soldiers lacked the imagination to exploit the disarray it caused. As events unfolded thereafter, gas, always unwelcome, never achieved the same element of surprise. It played only a marginal role in the outcome of most battles, with some exceptions such as Caporetto, where gas dislodged the Italians from their defensive positions in the mountains and enabled the Austrians to march into Italy.

The Poisonous Cloud is an outstanding history of chemical warfare in the First World War. Using previously unavailable archive material from Britain, France and Germany, Haber has written a brilliant analysis of the successive use of chlorine,

phosgene, mustard gas and arsenicals. Quoting figures of industrial output, he shows how the initial German superiority in gas was quickly eroded by the Allies. On both sides, however, there was muddle, and much bureaucratic wrangling over the type of gas to be used, the analysis of the effect of each of them, and whether it should be released from cylinders or shells.

According to the author, gas accounted for only some 3–3½ per cent of the casualties (dead and injured) of the war; this, it needs to be remembered, still represents over a million soldiers. Had the war continued unabated into 1919, Haber argues, the gas casualties, German in particular, would have been much higher because by

penetrating account. As for the Second World War, Spiers omits mention of the one serious incident, in the southern Italian town of Bari, which might have been the flashpoint for a further outbreak of chemical warfare. On 2 December 1943, the US ship *John Harvey* with its cargo of mustard gas was sunk in the aftermath of an air raid. The cargo was top secret and doctors had no knowledge of it, with a result that in the first 24 hours treatment of the casualties was inappropriate and led to the deaths of servicemen from many countries. Fortunately the Germans never cried foul, and the rules of engagement remained the same.

Spiers's account of the so called "yellow rain" incidents, where the US State De-

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

The blind lead the blind — casualties of a gas attack after the Battle of Estaires, 1918.

mid-1918 the Allied chemical warfare effort was in top gear and gas was to be used on an ever-increasing scale. Had that happened the interpretation of the value of gas in the First World War might have been different, but as things were Haber considers that it had only a marginal role on the outcome. Should chemicals be used in war today, the situation could well be the same because soldiers are now provided with a high degree of protection. Civilian casualties would be severe, however; simulation exercises of a chemical war in Europe, in which nerve gases are released, suggest that civilian casualties could outnumber the military by 20–1.

The Poisonous Cloud provides no clues as to what Fritz Haber would have thought of this outcome. More about the man himself and what drove him would have been welcome. His son says that as a result of the research for *The Poisonous Cloud* his admiration for his father "as an organiser and military thinker" has grown, and he is convinced that had the Germans not used gas in April 1915, someone else would have. Altogether the researching and writing of this book would seem to have been in part a catharsis for him.

By contrast it is difficult to find much to praise in Edward Spiers's *Chemical Warfare*. Little of it is new. As a history of the subject it is disappointing and the chapter dealing with the First World War is overshadowed by Haber's thorough and

partment has alleged that Soviet-backed forces have used mycotoxins as chemical warfare agents in south-east Asia and Afghanistan, gives the appearance of objectivity. In essence, however, it is a re-statement of the State Department case. Most scientists, and particularly those in government chemical warfare services, no longer believe the allegations and incline to the view that yellow rain is nothing other than bee excrement; even the State Department appears to be backing off on this subject, judging from its recent press releases.

Finally, Spiers advocates modernization of US chemical munitions in Germany. He believes that a comprehensive treaty on chemical weapons, with sufficient provisions for policing violations, will only be hammered out in Geneva — where negotiations for this have been in progress for ten years, and where much has already been agreed — if the United States has a better chemical arsenal. He may be right. But if he is wrong, and the US modernization programme for binary chemical weapons (agreed by Congress in December 1985) leads to a renewed chemical arms race, we will be even less secure. In the next chemical war it won't be the infantry that suffers, but the rest of us. □

Alastair Hay is a Lecturer in the Department of Chemical Pathology, University of Leeds, Leeds LS2 9JT, UK.